

New Jersey Institute of Technology  
Computer Science Department  
April 29, 2002

# Issues in the visualization and navigation of biomedical knowledge



*Olivier Bodenreider*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA

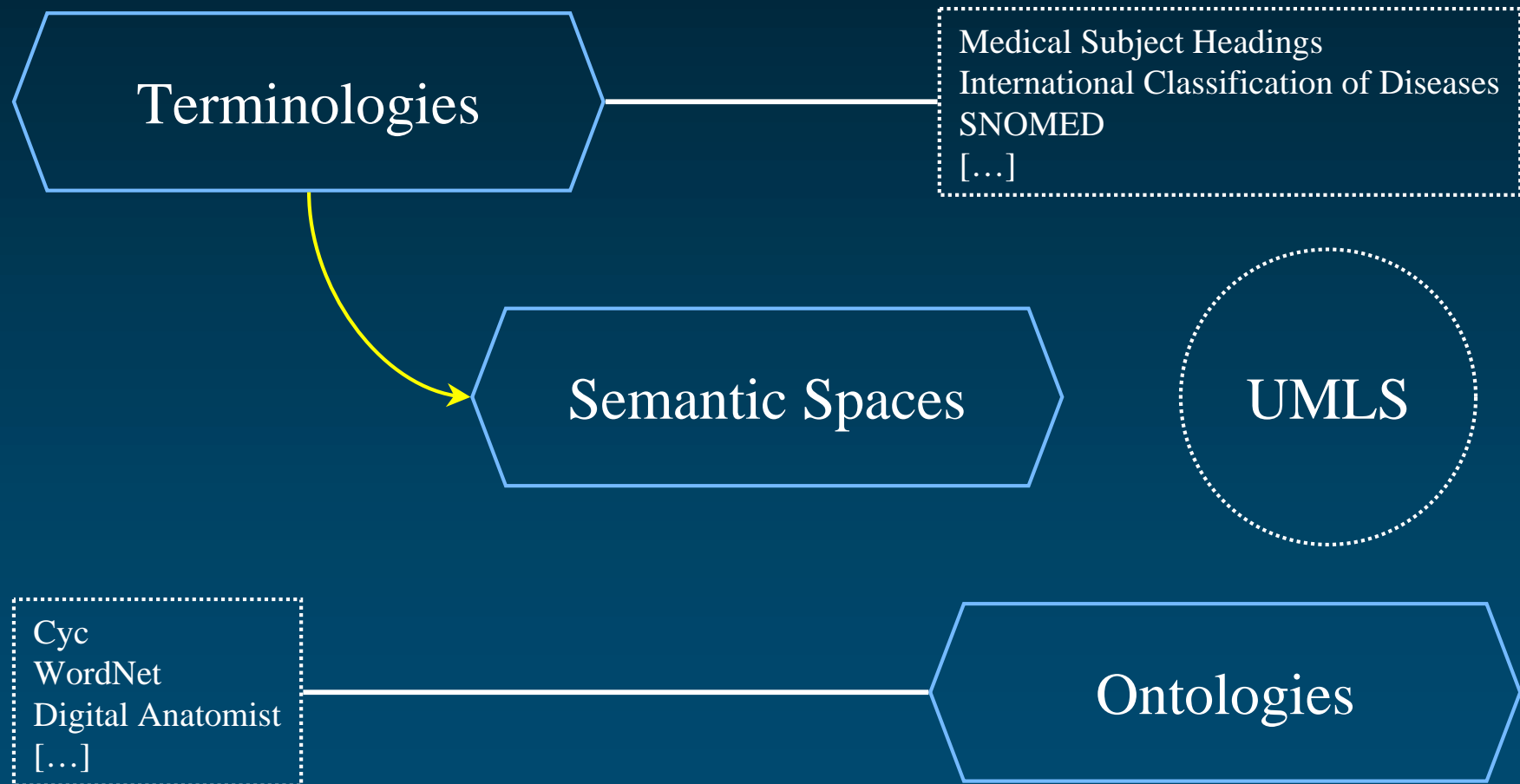
# Outline

---

- ◆ Biomedical terminologies as a source of biomedical knowledge
- ◆ Structural perspective on the Metathesaurus
- ◆ Visualizing biomedical knowledge
- ◆ From structure to semantics
  - Inherit relationships
  - Path between two concepts
  - Limitations

# Biomedical terminologies

# Biomedical knowledge organization



# Biomedical terminologies

---

## ◆ Core vocabularies

- anatomy (UWDA, Neuronames)
- drugs (First DataBank, Micromedex)
- medical devices (UMD, SPN)

## ◆ Several perspectives

- clinical terms (SNOMED, CTV3)
- information sciences (MeSH, CRISP)
- administrative terminologies (ICD-9-CM, CPT-4)
- standards (HL7, LOINC)



# Biomedical terminologies (cont'd)

---

- ◆ Specialized vocabularies
  - nursing (NIC, NOC, NANDA, Omaha, PCDS)
  - dentistry (CDT)
  - oncology (PDQ)
  - psychiatry (DSM, APA)
  - adverse reactions (COSTART, WHO ART)
  - primary care (ICPC)
- ◆ Knowledge bases (AI/Rheum, DXplain, QMR)

# UMLS

## ◆ Two-level structure

- Semantic Network

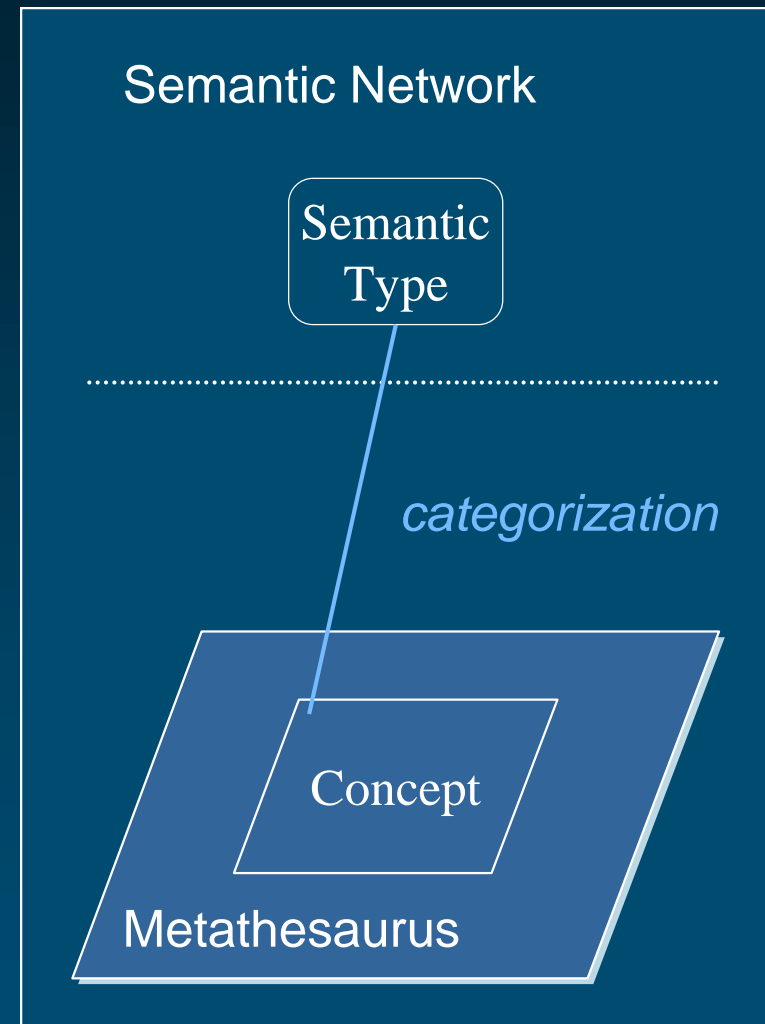
- 134 Semantic Types (STs)
- Relationships among STs

- Metathesaurus

- 800,000 concepts
- Inter-concept relationships

- Link = categorization

- Often isa
- Rarely is an instance of



## Semantic Types

Anatomical  
Structure

Fully Formed  
Anatomical  
Structure

Embryonic  
Structure

Disease or  
Syndrome

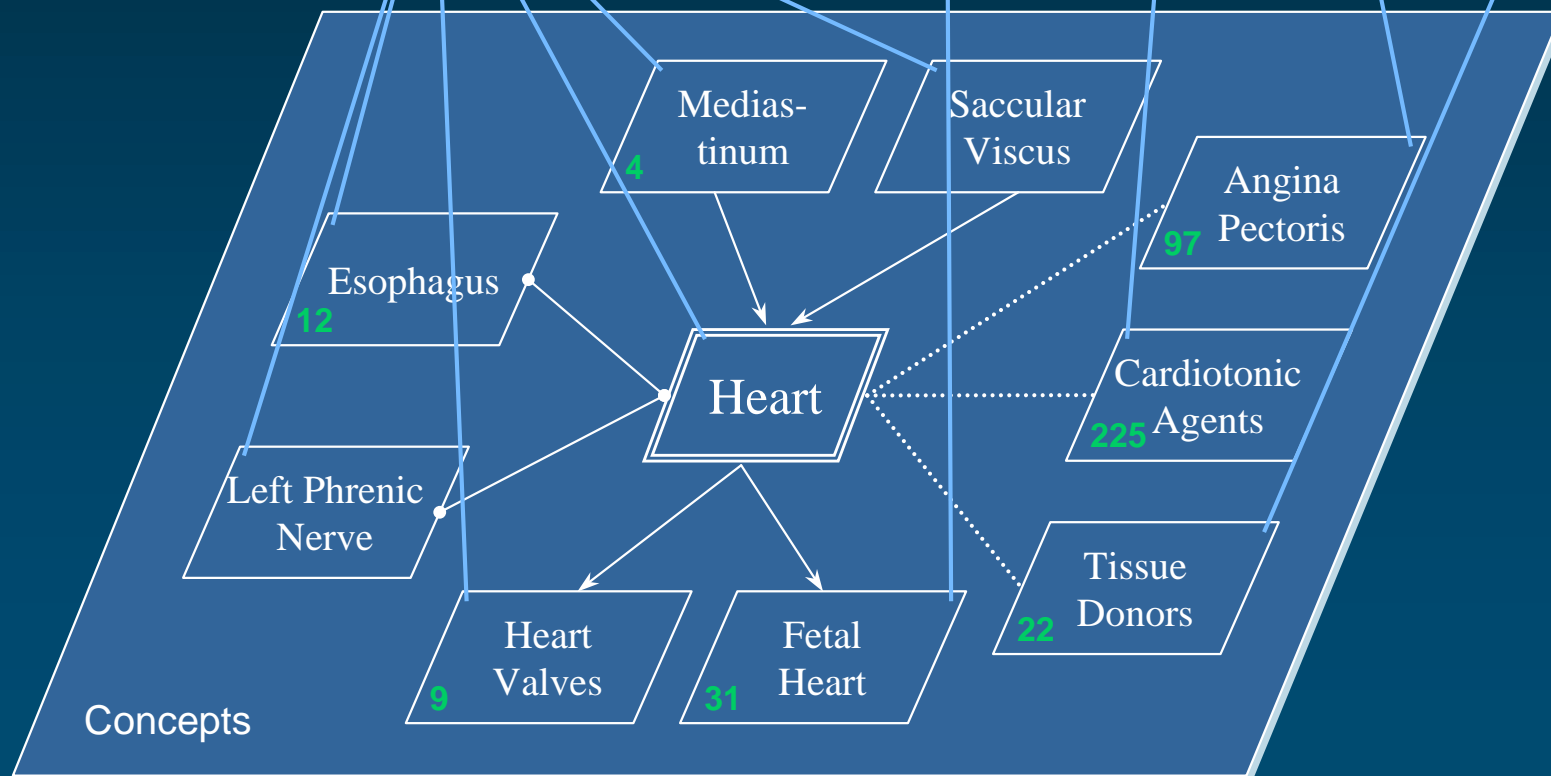
Body Part, Organ or  
Organ Component

Pharmacologic  
Substance

Population  
Group

*Semantic  
Network*

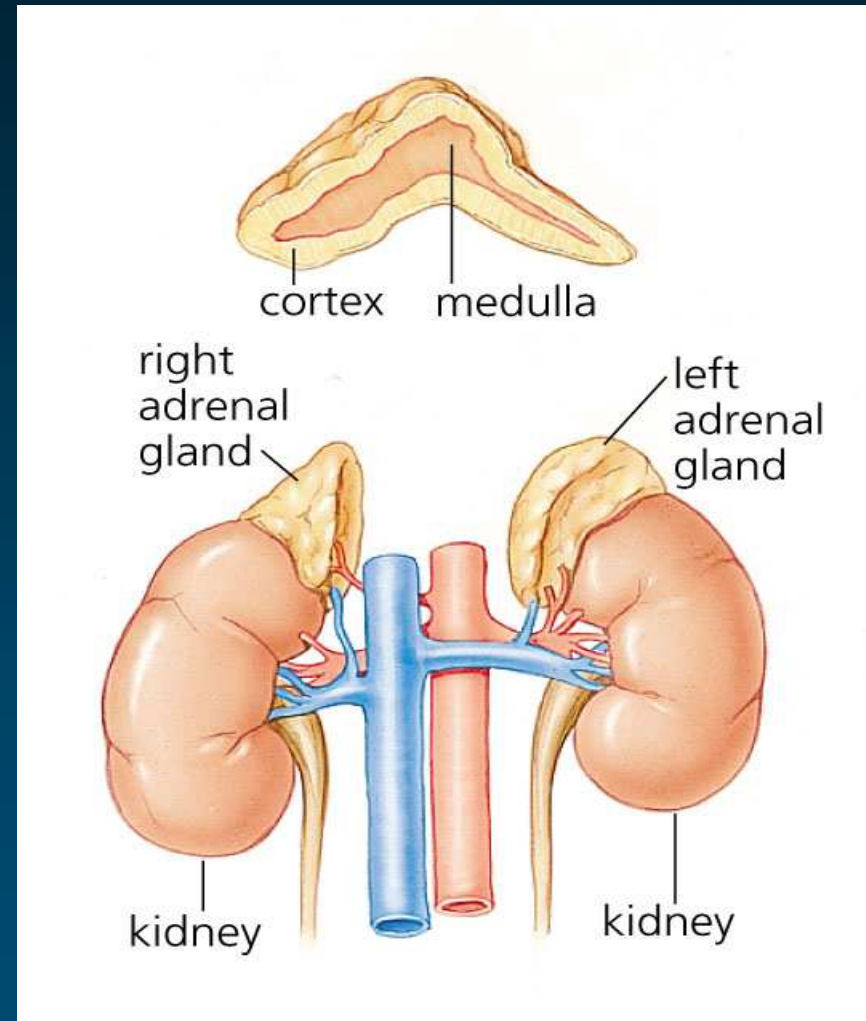
*Metathesaurus*





# Addison's disease

- ◆ Addison's disease is a rare endocrine disorder
- ◆ Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
- ◆ For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism



# Structural perspective on the Metathesaurus

# Hierarchy

## ◆ Hierarchical relationships

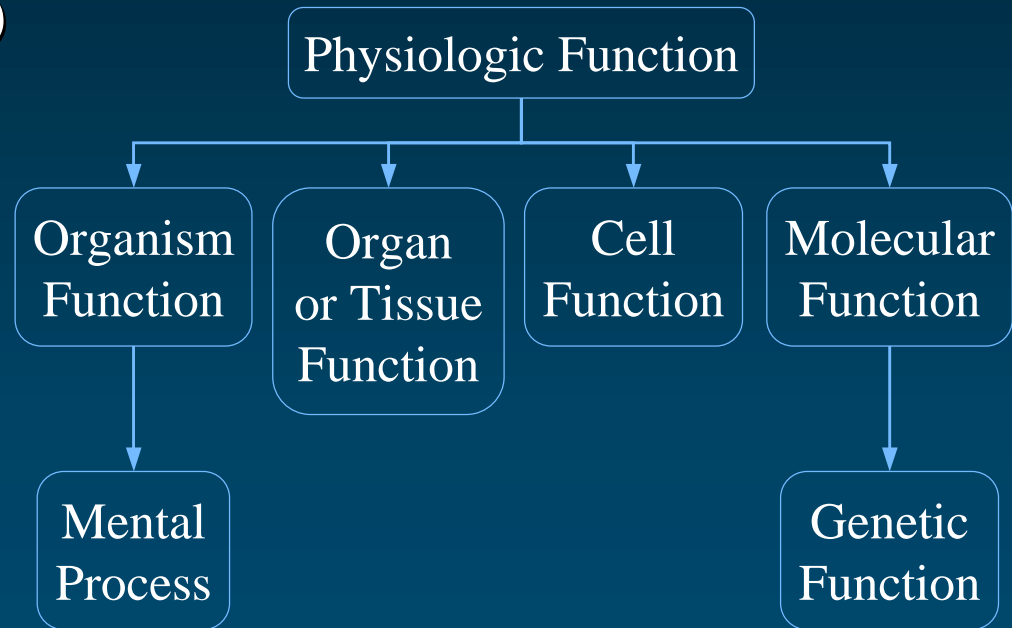
- Taxonomy (isa)
- Meronymy (part of)

## ◆ Partial ordering

- [Reflexivity]
- Antisymmetry
- Transitivity

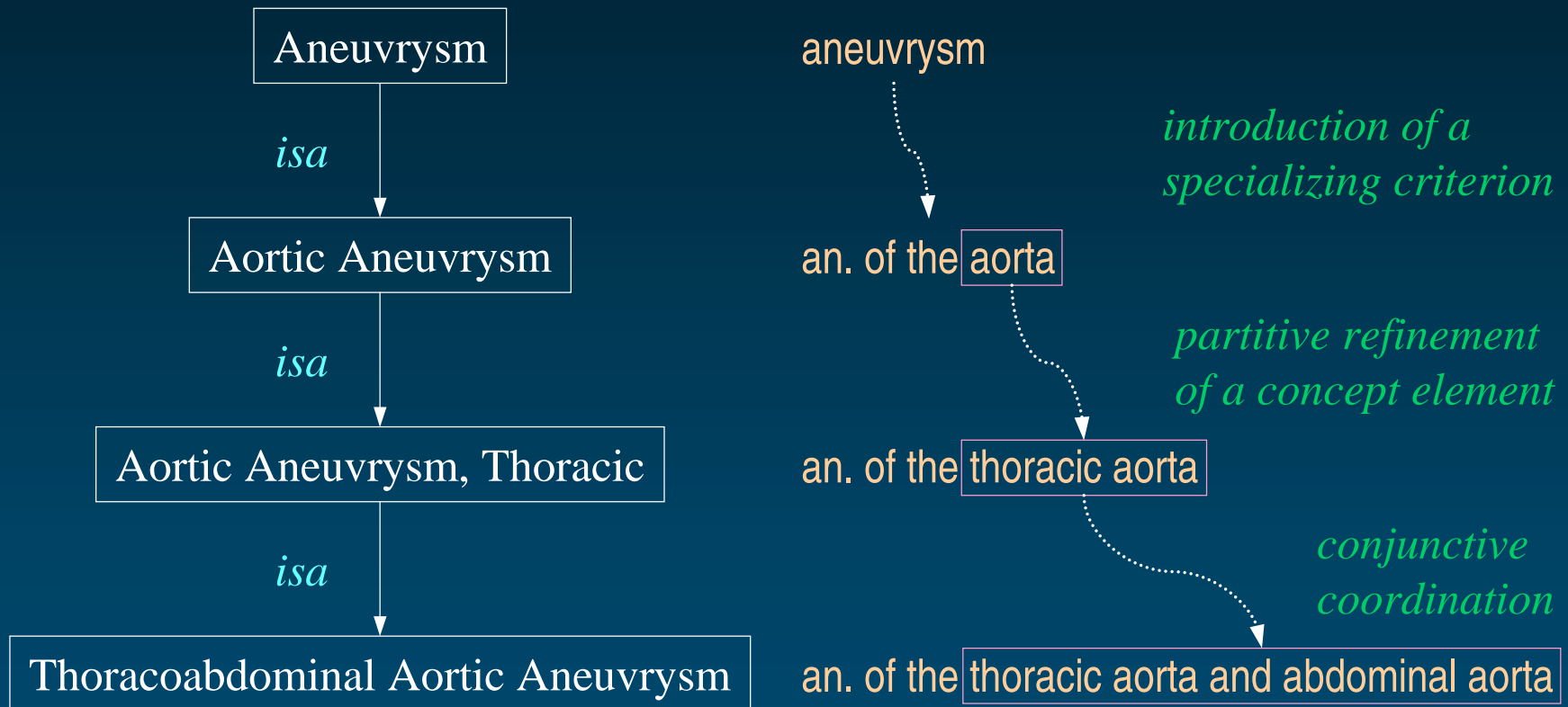
## ◆ Inheritance

## ◆ Reasoning



# Principles of subsumption

[Bernauer, AMIA 1994]



# Hierarchies in source vocabularies

---

## ◆ Structure

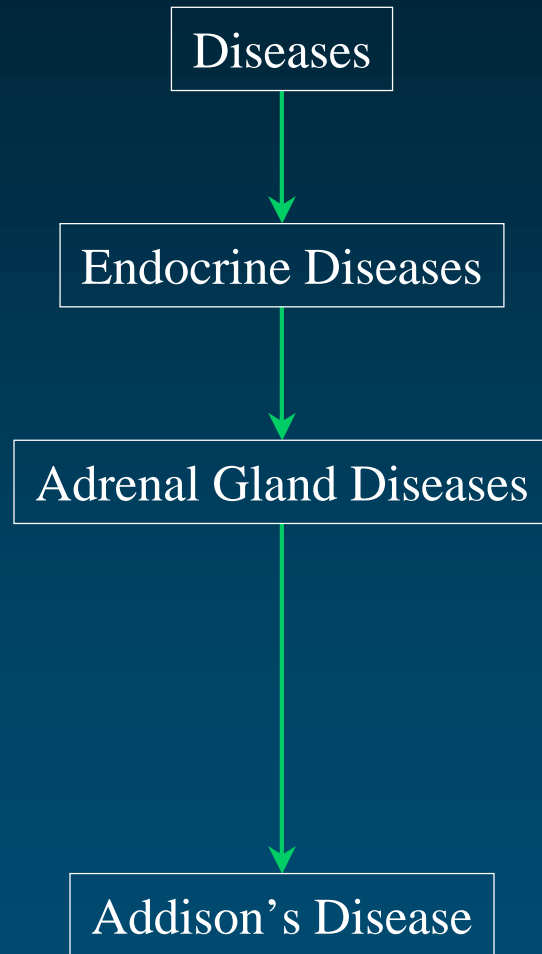
- Single tree
- Polyhierarchical (multiple parents)

## ◆ Relationships

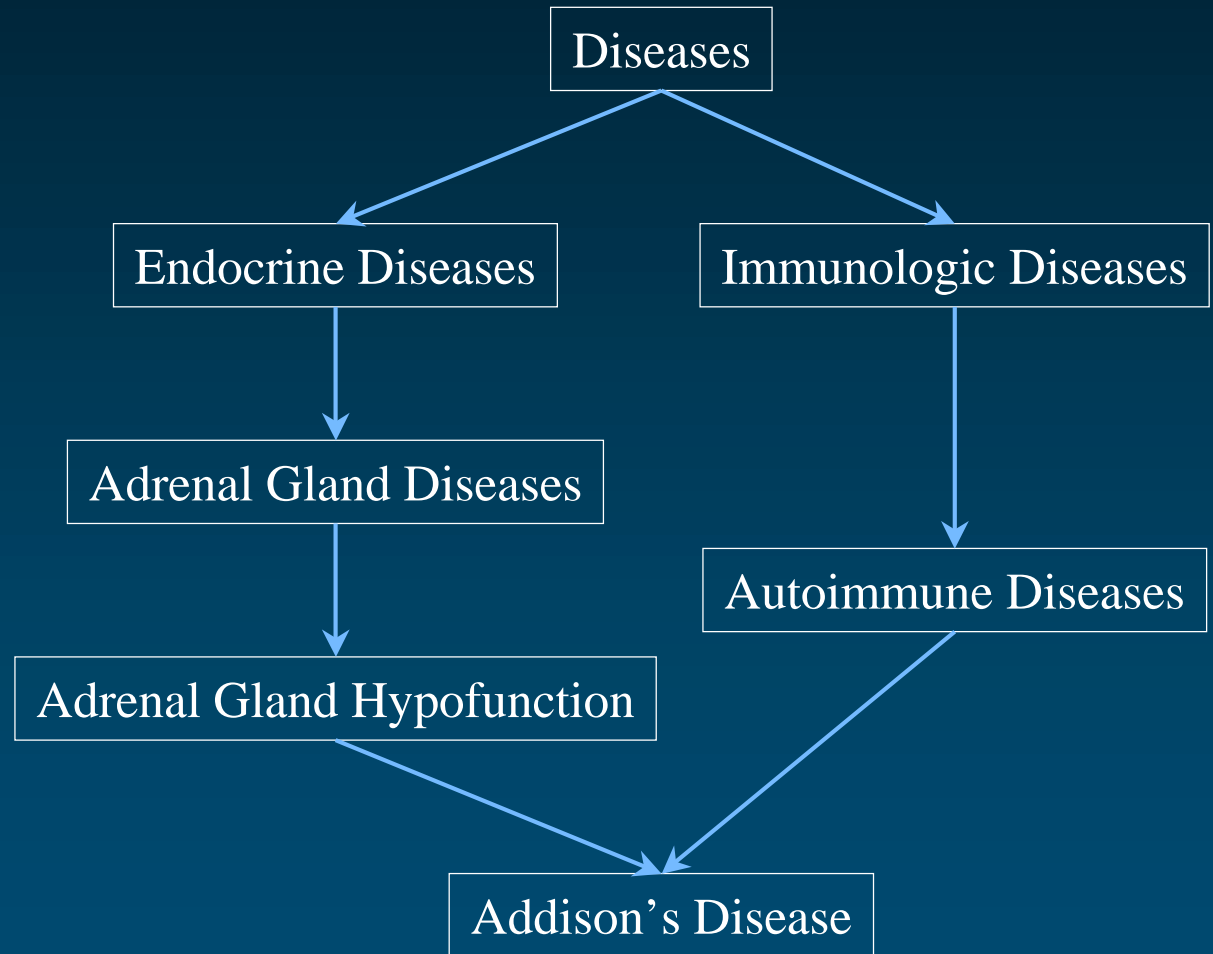
- Usually implicit
- May be other than *isa* or *part of*
  - E.g., Thesaurus relationships

# AD in medical vocabularies Contexts

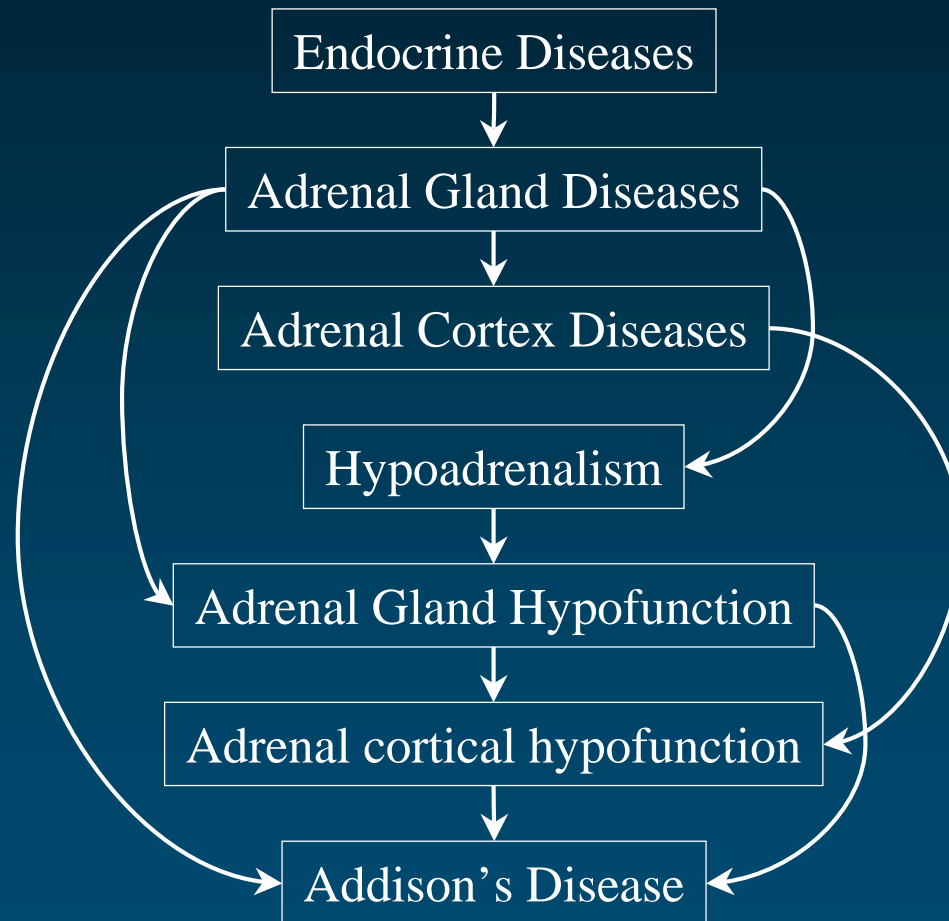
## SNOMED



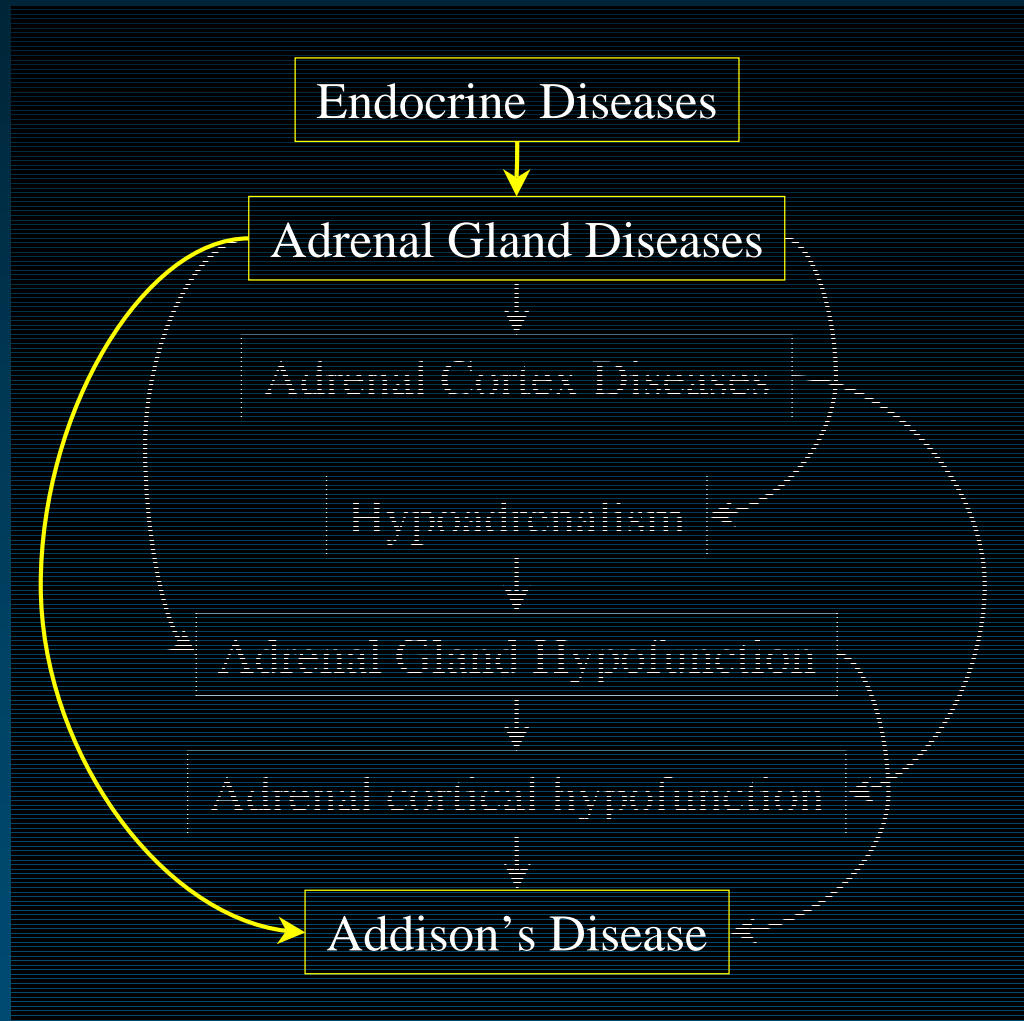
## MeSH



# AD in UMLS Contexts

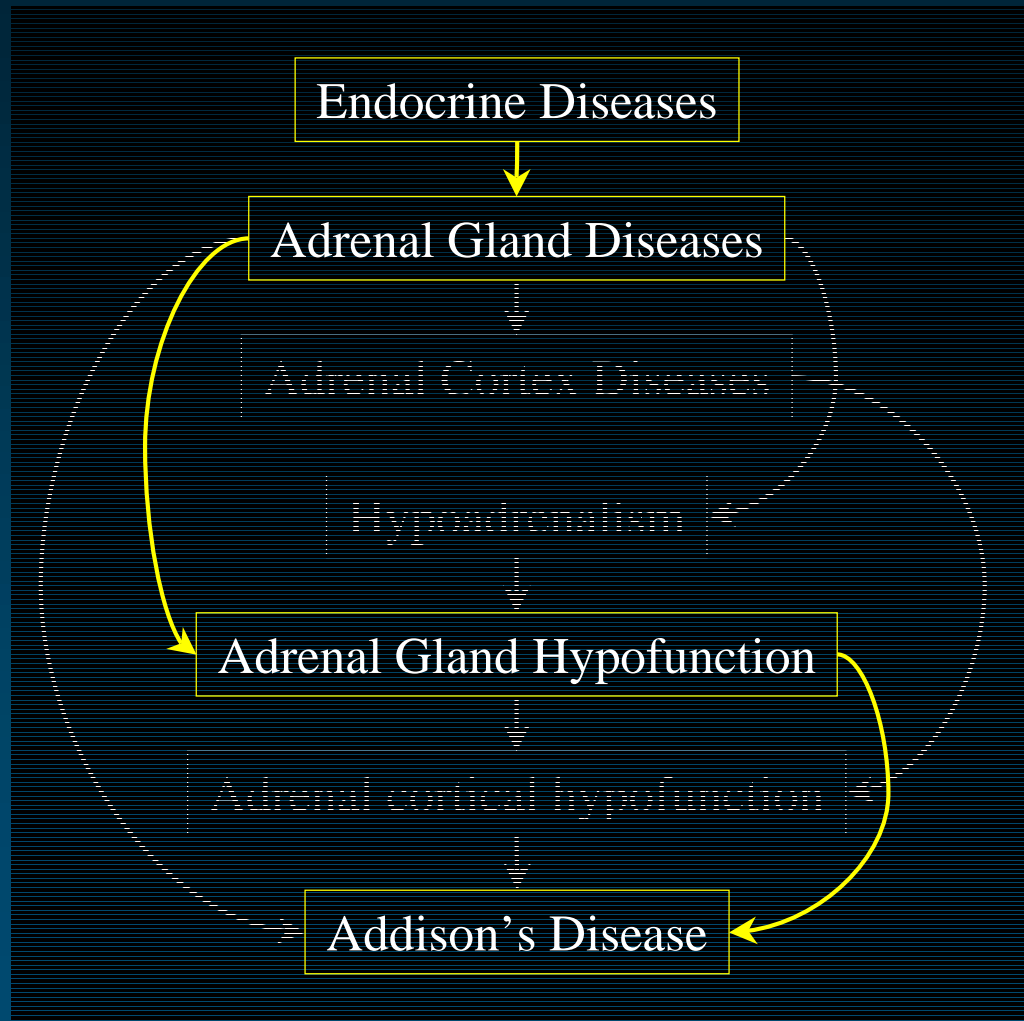


# AD in UMLS SNOMED context

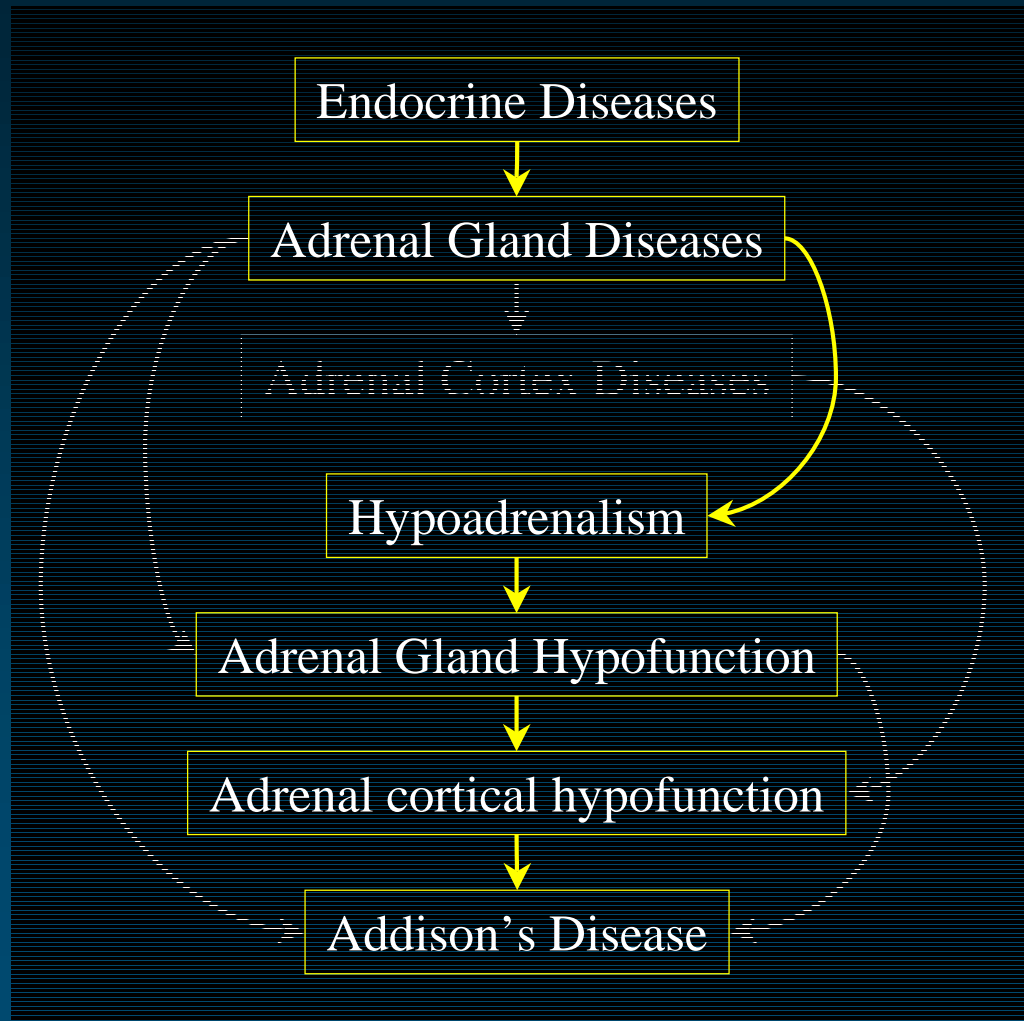




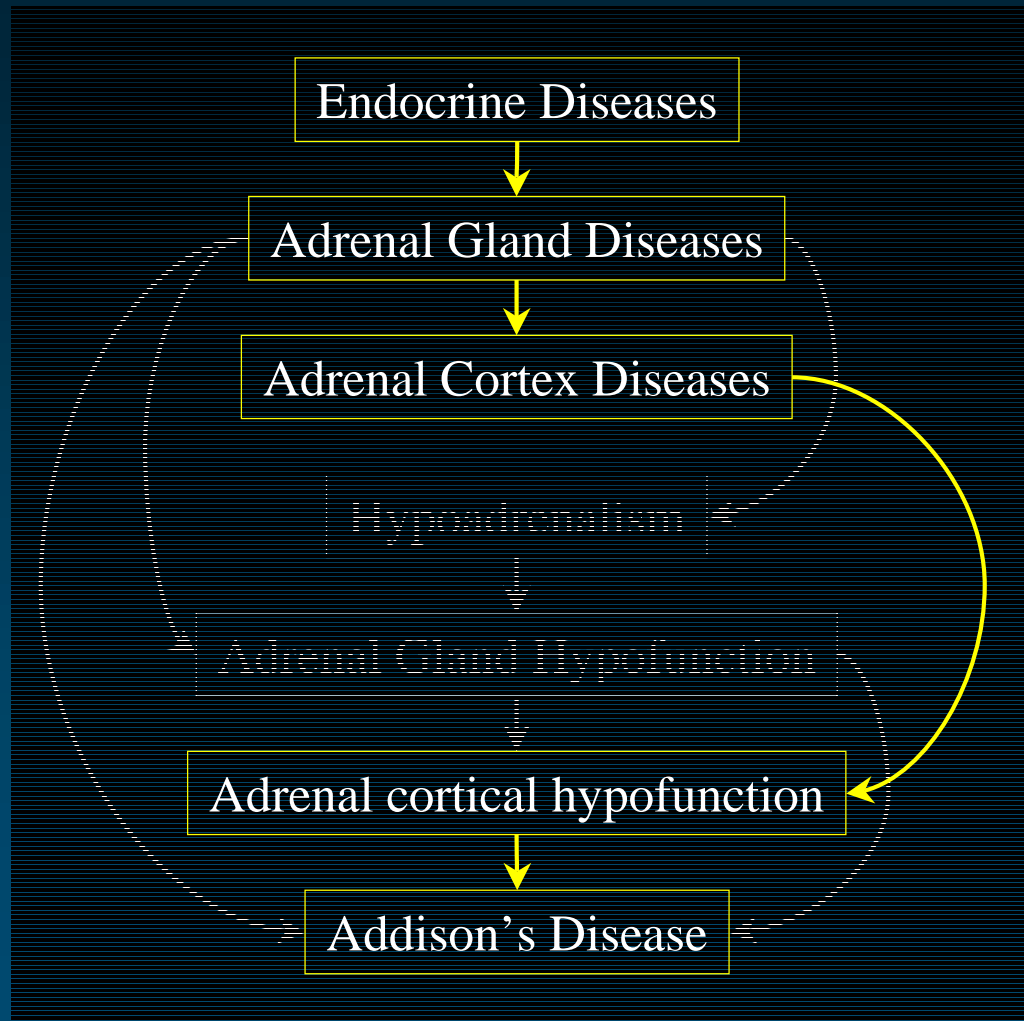
# AD in UMLS MeSH context

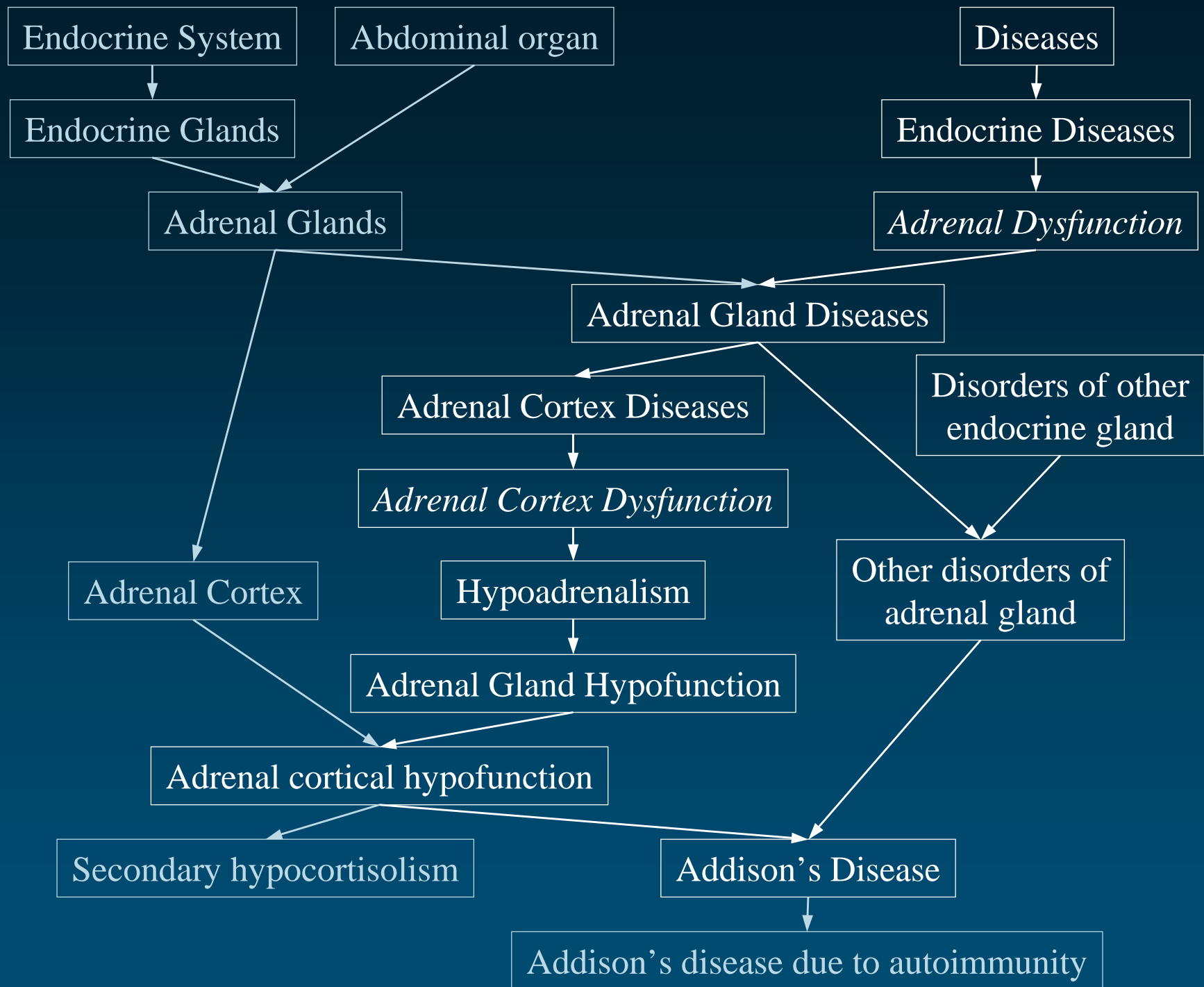


# AD in UMLS Read Codes context



# AD in UMLS AOD Thes. context





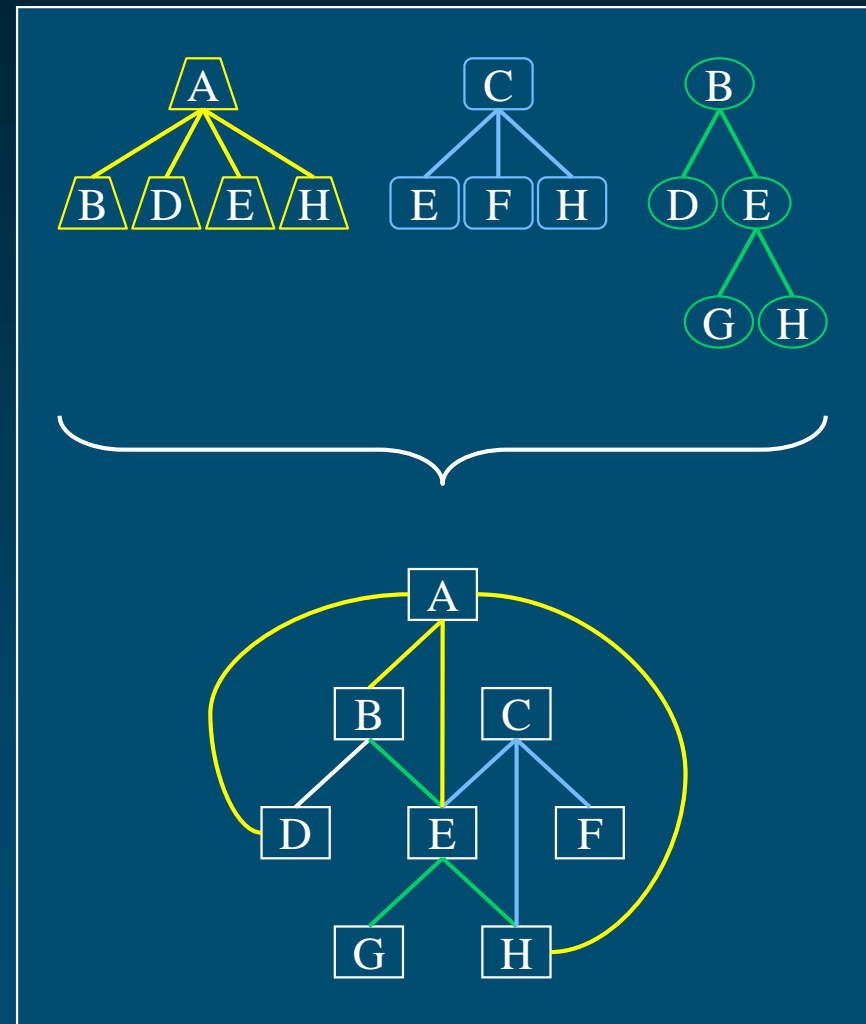
# Hierarchies in source vocabularies

---

- ◆ Often task-driven  
rather than based on principles
- ◆ Usually suitable for information retrieval
  - Better recall
  - Precision may not be crucial
- ◆ Not necessarily suitable for reasoning
- ◆ But expected to be consistent structurally

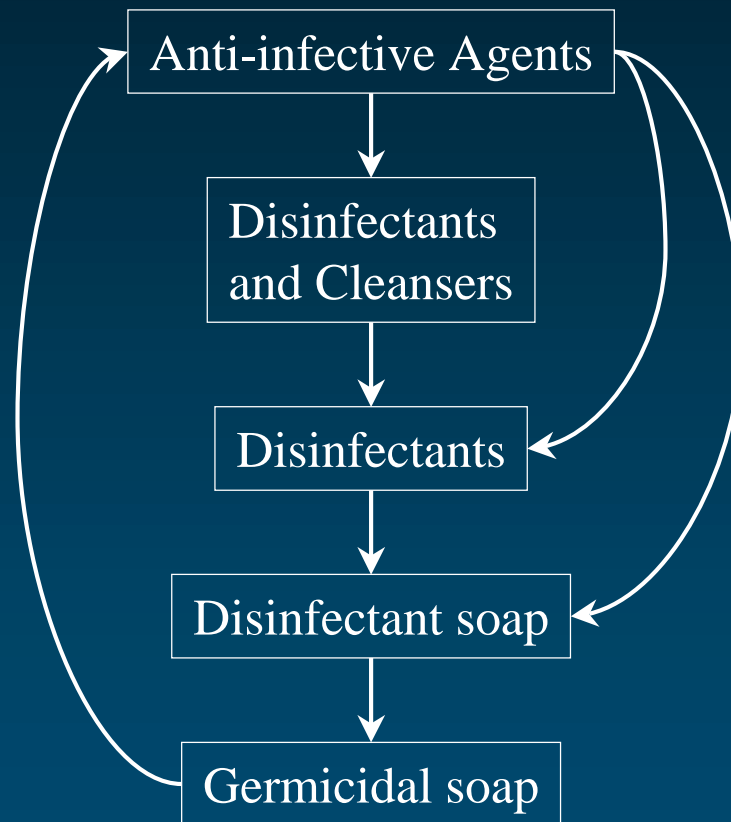
# AD in UMLS Contexts

- ◆ Multiple **tree** structures combined into a **graph** structure
- ◆ Directed **acyclic** graph (DAG)



# Actually, there are some cycles

---



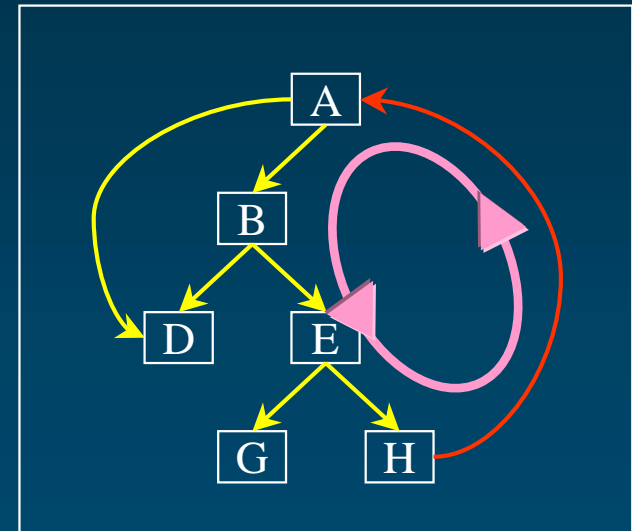
# Issues with cycles

## ◆ Theoretical

- Violate the antisymmetry property of partial ordering relations

## ◆ Practical

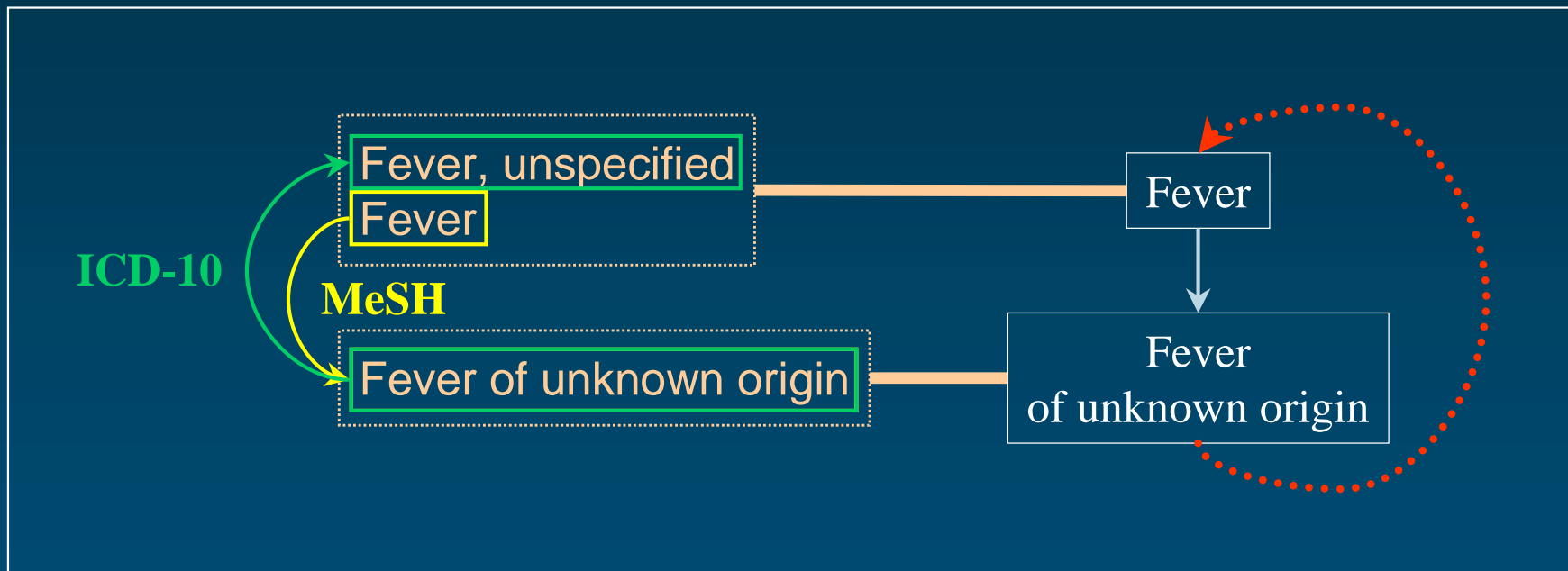
- Loops in graph traversal
- Impossible to perform transitive reduction





# Cycle due to underspecification

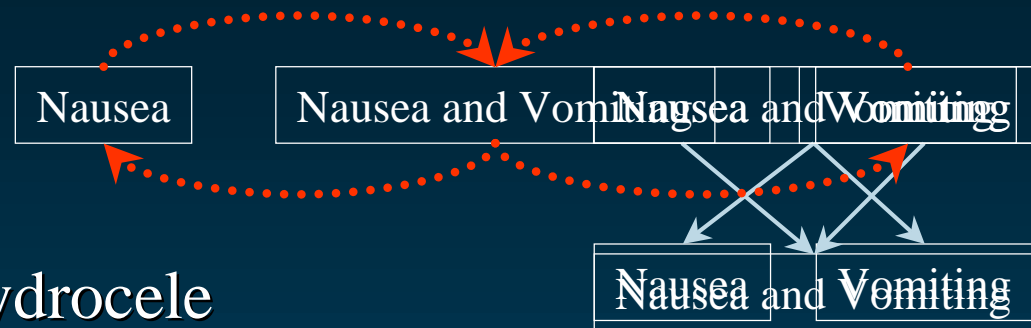
- ◆ Specified and underspecified terms
  - May appear at different levels in a source hierarchy
  - Are clustered into the same concept (same meaning)



# Other causes

[Bodenreider, AMIA 2001]

- ◆ Compound terms
  - HYDROCELE, Hydrocele
- ◆ Metadata
- ◆ Classes and member
  - Purines, Purine
- ◆ Organizational conventions
  - Acid + Base  $\rightleftharpoons$  Salt + Water
- ◆ Idiopathic
  - Wrong relationships
  - Use of non-hierarchical relationships in “hierarchies”



# Visualizing biomedical knowledge

# Visualizing biomedical knowledge

---

## ◆ Objectives

- Make knowledge navigable by users
- Make knowledge available to applications

## ◆ Common issues

- Reduce complexity
- Provide consistent views across the domain
- Extend views to fit specific needs

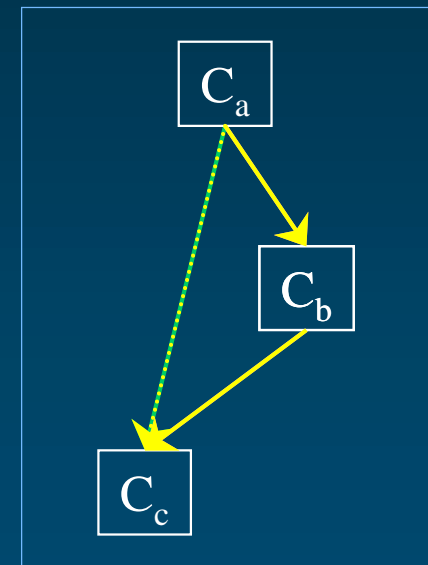
## ■ *UMLS Semantic Navigator*

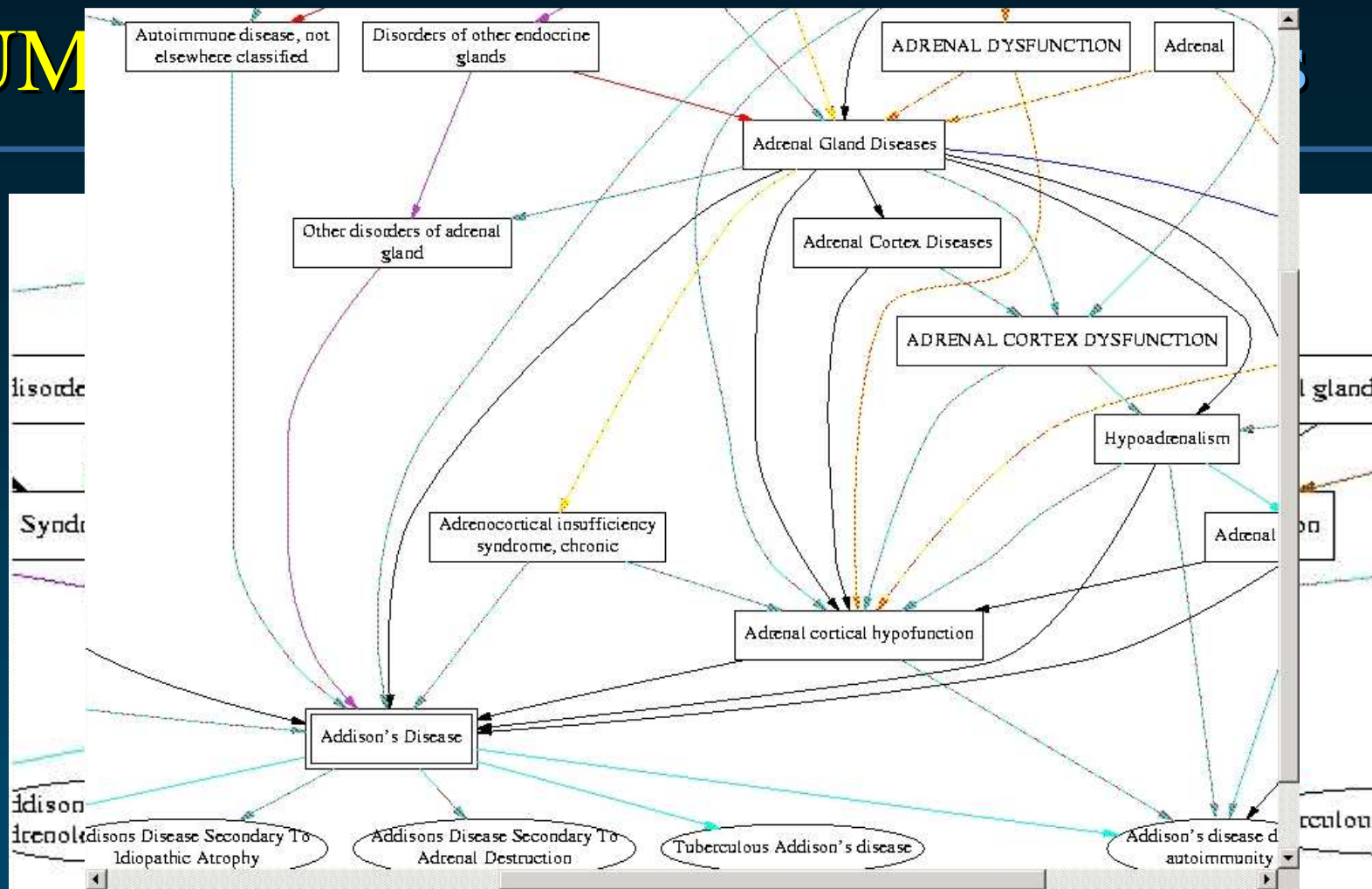
[umlsks.nlm.nih.gov](http://umlsks.nlm.nih.gov) → Resources → Semantic Navigator



# UMLS Semantic Navigator

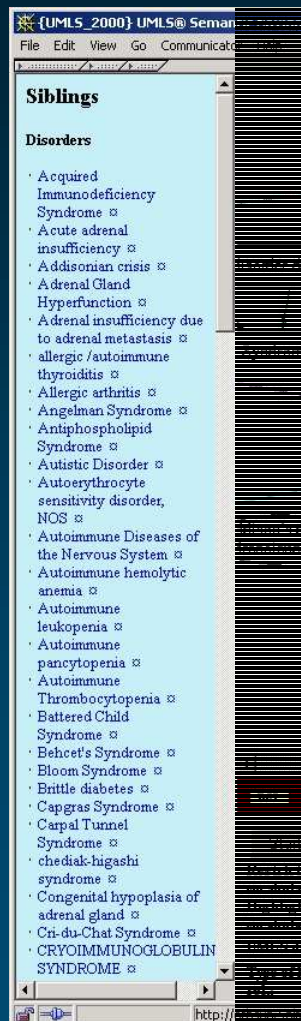
- ◆ Visualize semantic locality
- ◆ Features
  - All relationships presented simultaneously
    - Metathesaurus relationships
    - Semantic network relationships
  - Hierarchical relationships presented graphically
  - Dynamic and navigable
  - Transitive reduction



[illegible]

# UMLS Semantic

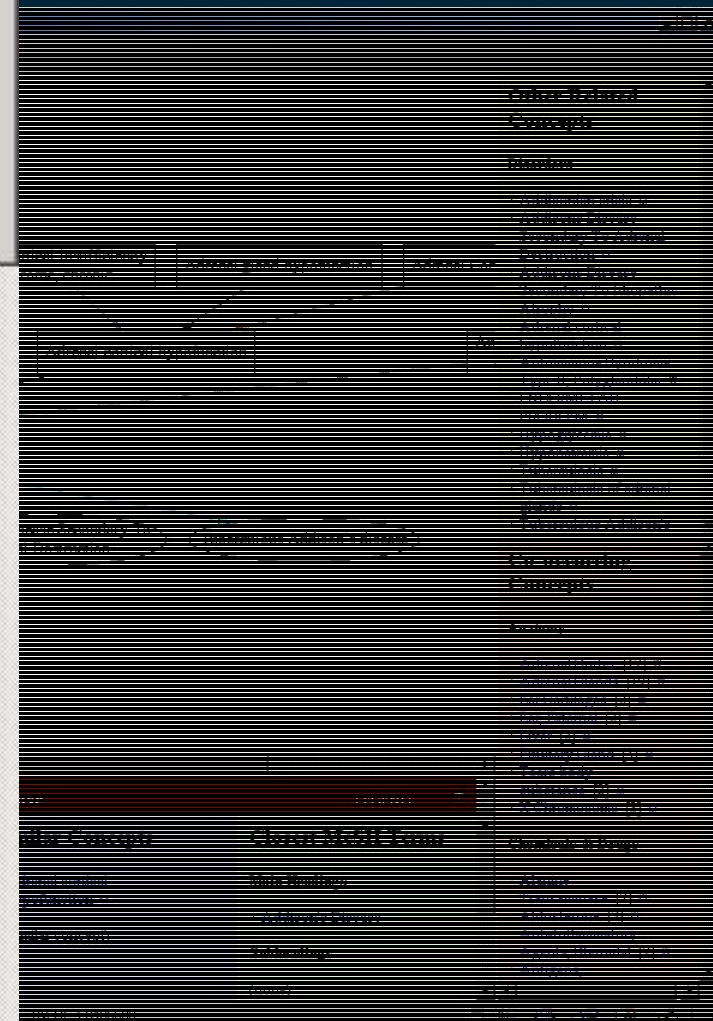
# Investigator Concepts



## Siblings

### Disorders

- Acquired Immunodeficiency Syndrome ☒
- Acute adrenal insufficiency ☒
- Addisonian crisis ☒
- Adrenal Gland Hyperfunction ☒
- Adrenal insufficiency due to adrenal metastasis ☒
- allergic/autoimmune thyroiditis ☒
- Allergic arthritis ☒
- Angelman Syndrome ☒
- Antiphospholipid Syndrome ☒
- Autistic Disorder ☒
- Autoerythrocyte sensitivity disorder, NOS ☒
- Autoimmune Diseases of the Nervous System ☒
- Autoimmune hemolytic anemia ☒
- Autoimmune leukopenia ☒
- Autoimmune pancytopenia ☒
- Autoimmune Thrombocytopenia ☒
- Battered Child Syndrome ☒



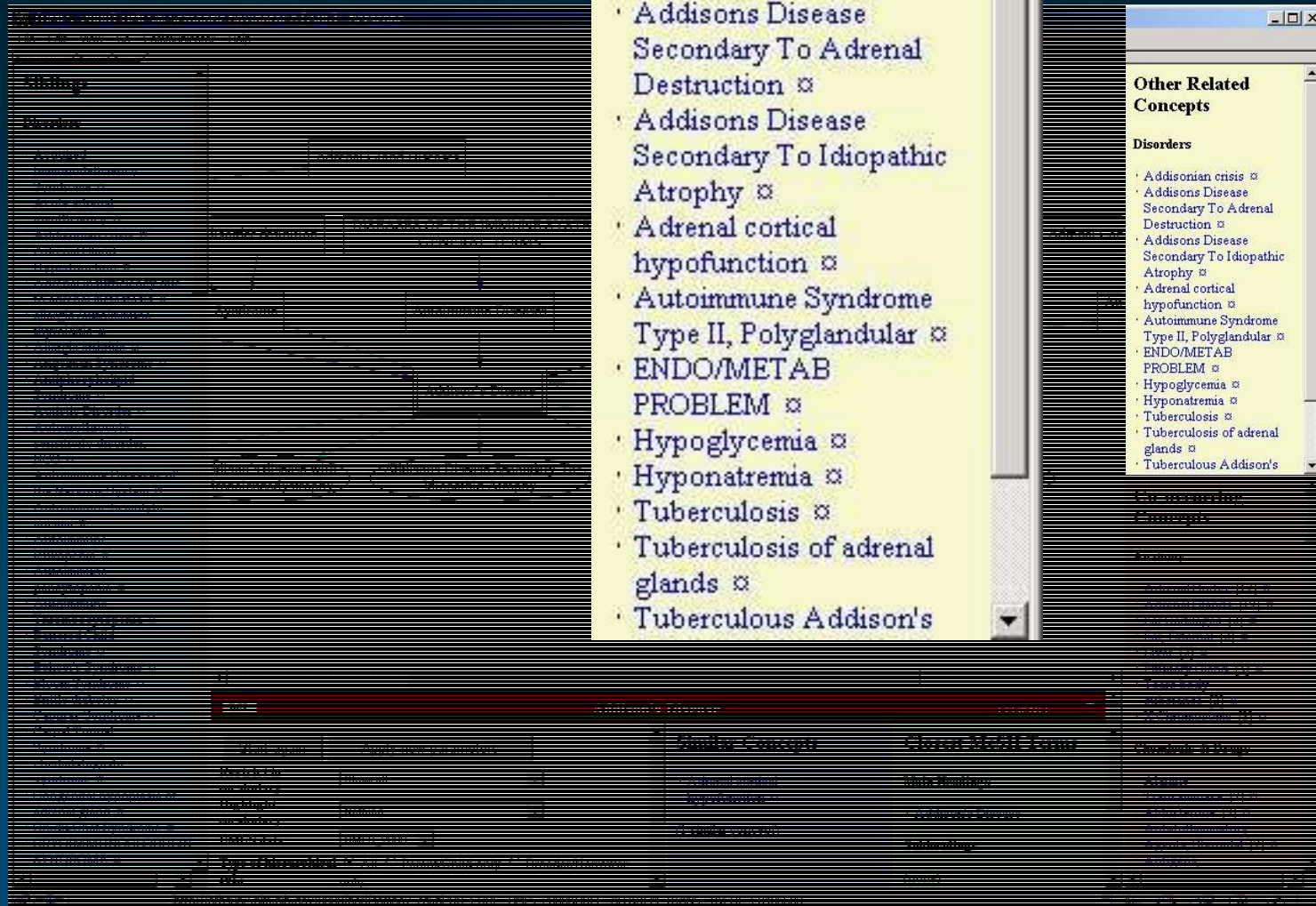


# UMLS Semantic Network

## Other Related Concepts

### Disorders

- Addisonian crisis ☐
- Addison's Disease
- Secondary To Adrenal Destruction ☐
- Addison's Disease
- Secondary To Idiopathic Atrophy ☐
- Adrenal cortical hypofunction ☐
- Autoimmune Syndrome Type II, Polyglandular ☐
- ENDO/METAB PROBLEM ☐
- Hypoglycemia ☐
- Hyponatremia ☐
- Tuberculosis ☐
- Tuberculosis of adrenal glands ☐
- Tuberculous Addison's





# UMLS Semantic Navigator Concepts

The screenshot displays the UMLS Semantic Navigator interface. On the left, a hierarchical concept map is visible, showing relationships between various medical concepts. The central pane, titled "Co-occurring Concepts", lists concepts under two main categories: "Anatomy" and "Chemicals & Drugs".

**Co-occurring Concepts**

**Anatomy**

- Adrenal Cortex [12] ☐
- Adrenal Glands [19] ☐
- Ear Cartilages [2] ☐
- Ear, External [2] ☐
- Liver [2] ☐
- Pituitary Gland [3] ☐
- Tears body substance [2] ☐
- X Chromosome [3] ☐

**Chemicals & Drugs**

- Alanine Transaminase [2] ☐
- Aldosterone [3] ☐
- Anti-Inflammatory Agents, Steroidal [2] ☐
- Antigens,

On the right side of the interface, there is a sidebar with additional concept lists, including "Other Related Concepts" and "Co-occurring Concepts" (repeated). The "Co-occurring Concepts" sidebar also lists concepts under "Anatomy" and "Chemicals & Drugs".

# UMLS Semantic Network

Relationships  
of **Addison's Disease** (C1)  
*Disease or Syndrome*  
to **Adrenal Cortex** (C2)  
*Body Part, Organ, or Organ Component*

---

**Metathesaurus Relationships**

C1 *co-occurs with* C2

Frequency = 12 • MEDLINE

---

**Semantic Network Relationships**

<i>Disease or Syndrome</i>	• <i>has_location</i>	<i>Body Part, Organ, or Organ Component</i>
----------------------------	-----------------------	---

[Close this window](#)

---

Interface version: 2.01 UMLS data: UMLS\_2000

Relationships  
of **Addison's Disease** (C1)  
*Disease or Syndrome*  
to **Adrenal Cortex** (C2)  
*Body Part, Organ, or Organ Component*

---

**Metathesaurus Relationships**

C1 *co-occurs with* C2

Frequency = 12 • MEDLINE

---

**Semantic Network Relationships**

<i>Disease or Syndrome</i>	• <i>has_location</i>	<i>Body Part, Organ, or Organ Component</i>
----------------------------	-----------------------	---

[Close this window](#)

---

Interface version: 2.01 UMLS data: UMLS\_2000



# Visualization Research directions

---

- ◆ Selecting relevant co-occurring concepts
  - Based on the relative frequency
  - Compared to symbolic relationships
- ◆ Visualizing the paths between two concepts
  - Display polyhierarchy
  - Graph theory algorithms

# Object-oriented model Specifications

---

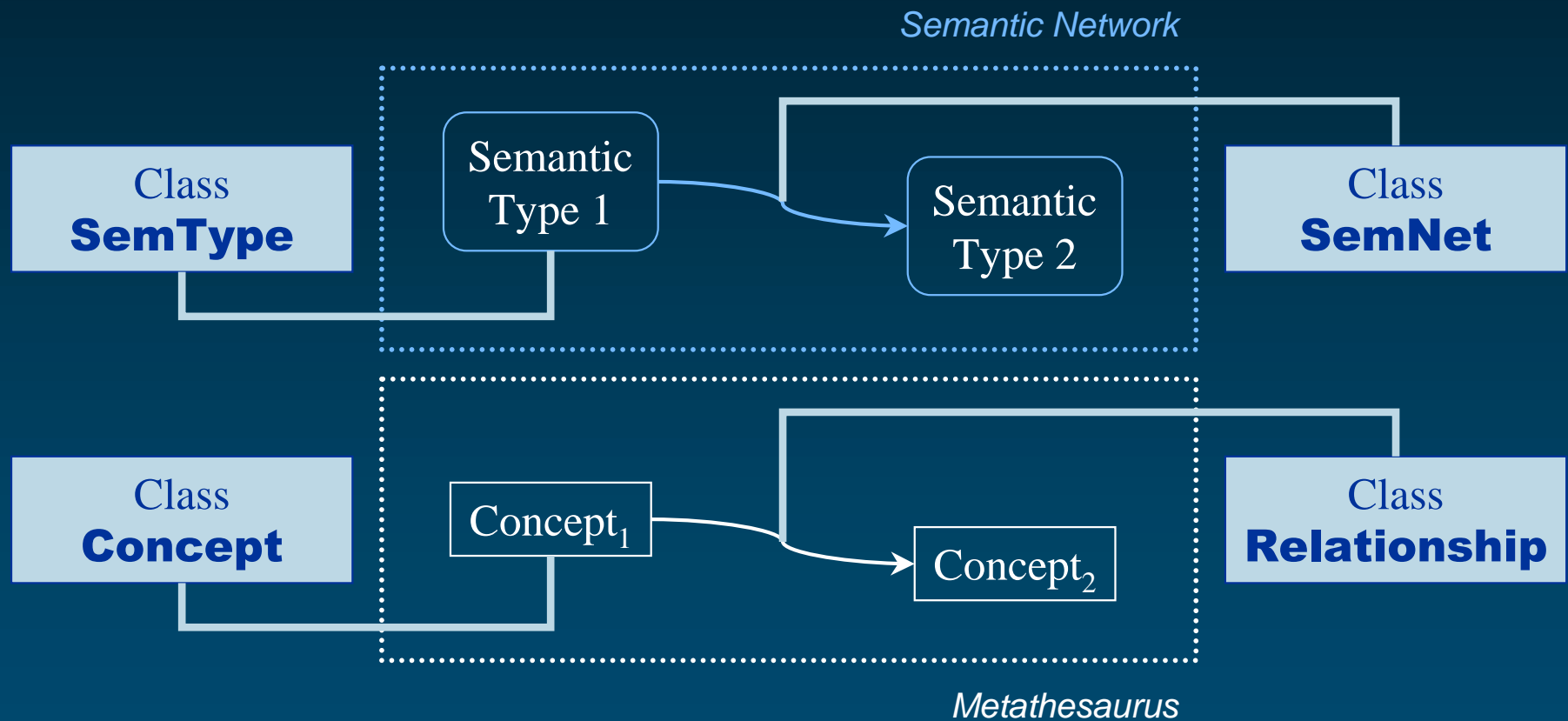
- ◆ Knowledge-oriented
- ◆ Simple
- ◆ High-level methods
- ◆ Independent
  - From the UMLS relational format
  - From back-end implementation
- ◆ Extendable

# Design choices

---

- ◆ Object-oriented model
  - Extensible through derivation
- ◆ Limited number of classes
  - Not as comprehensive as the UMLS itself
- ◆ Methods
  - Accessor methods (properties)
  - High-level methods

# 4 major classes



# Class **Concept**

---

## ◆ Properties

- Unique identifier (CUI)
- List of synonymous terms [in a given source/language]
- List of definitions
- List of sources
- Set of related concepts (instances of **Concept**)
- Set of semantic types (instances of **SemType**)
- *Total frequency of co-occurrence in MEDLINE*

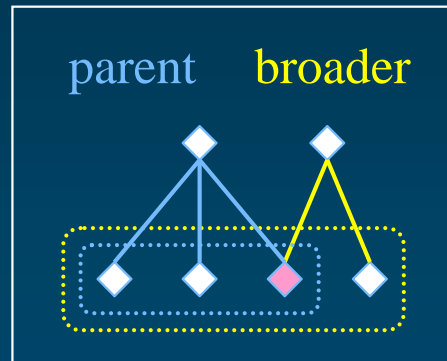
# Class Concept

- ◆ Methods provided for convenience

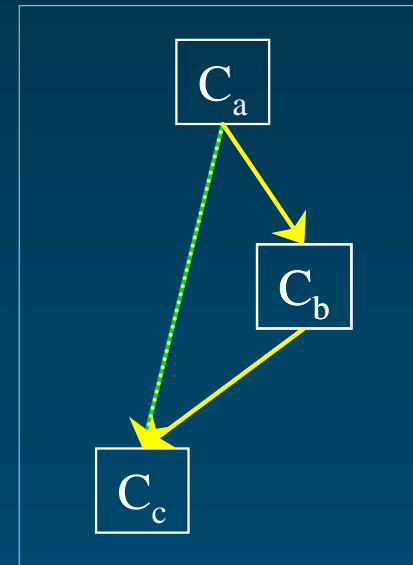
- $\text{anc1} = \text{par} \cup \text{bro}$

- ◆ Higher-level methods

- sibx: Extended siblings

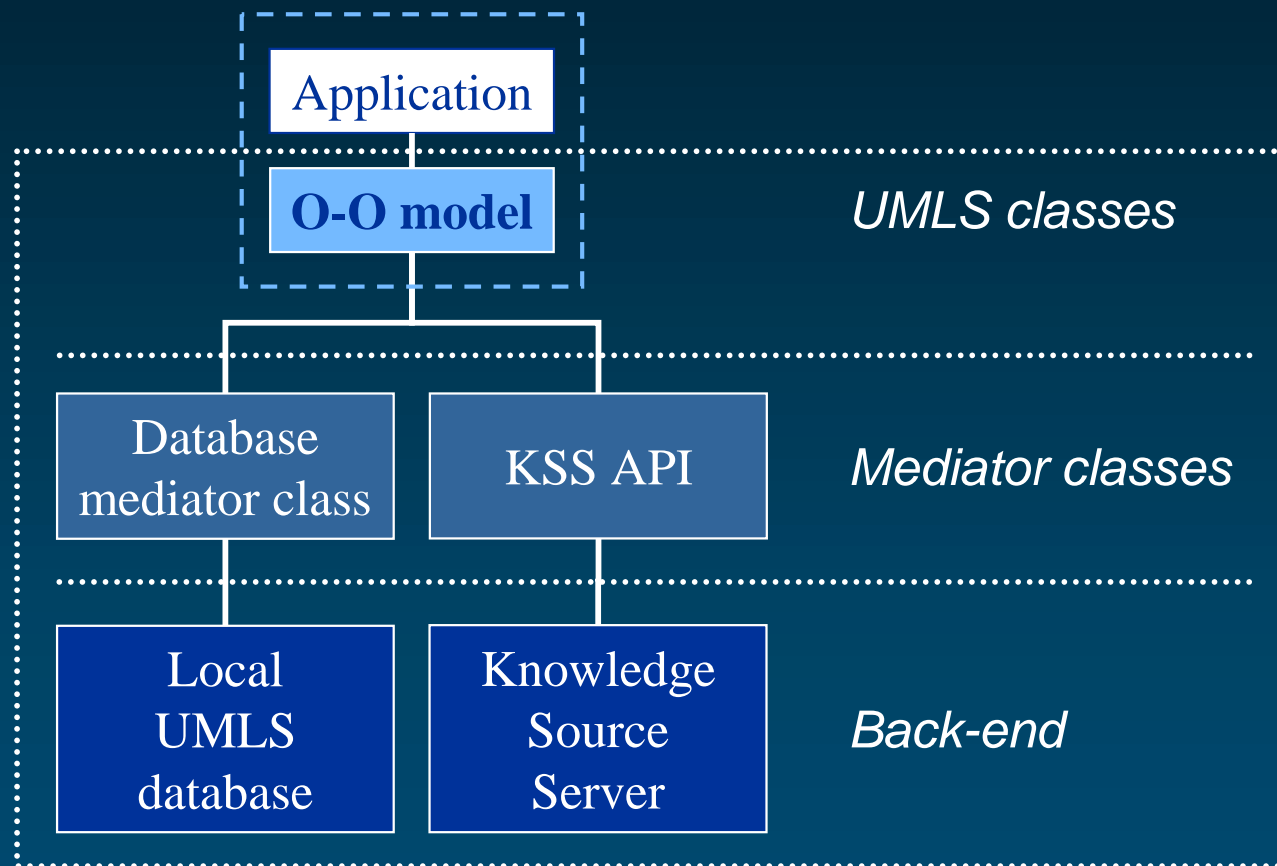


- par\_tr: Transitive reduction





# Architecture



# From structure to semantics

## *1. Inherit relationships*

# Semantic Network

---

## ◆ Semantic types (134)

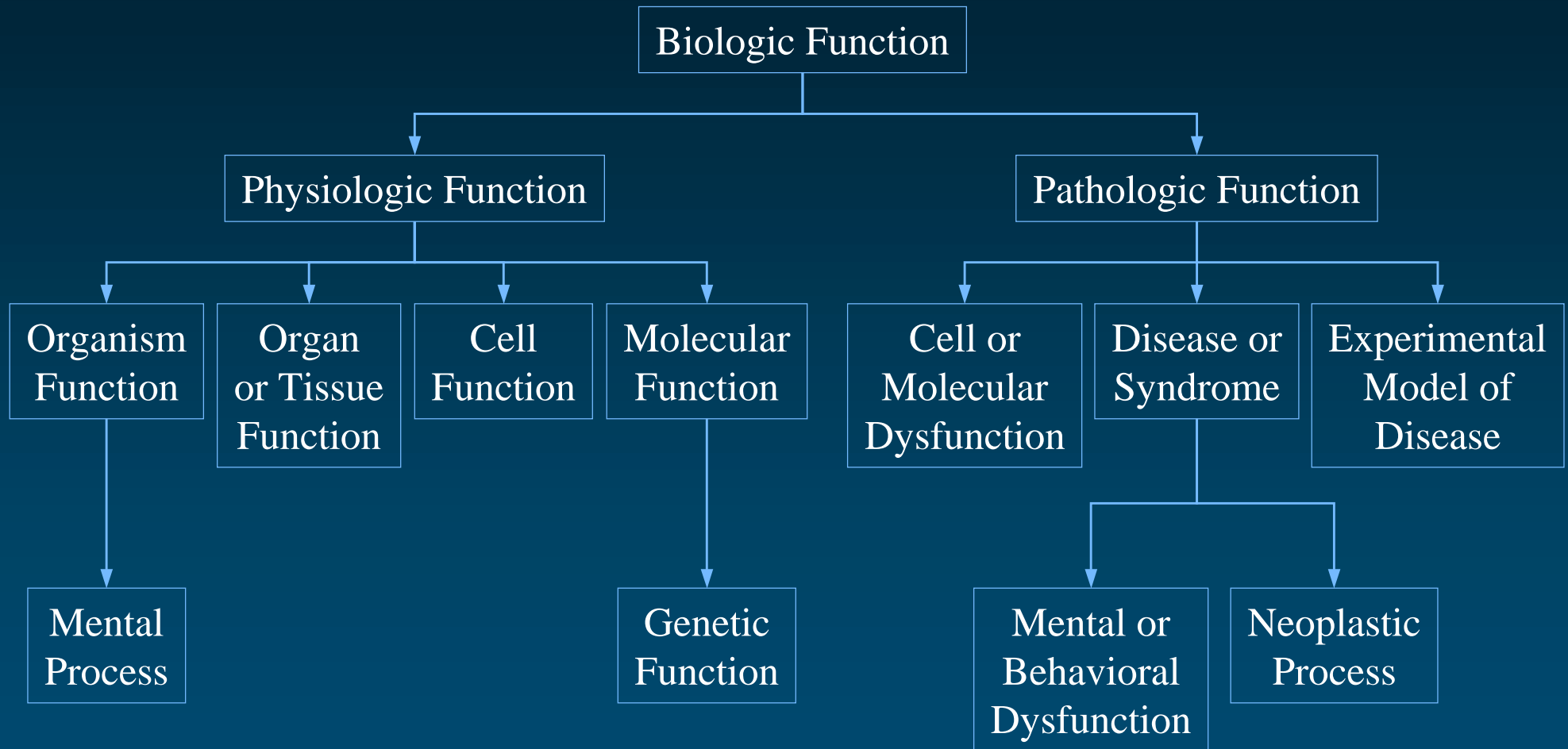
- tree structure
- 2 major hierarchies
  - Entity
    - Physical Object
    - Conceptual Entity
  - Event
    - Activity
    - Phenomenon or Process

# Semantic Network

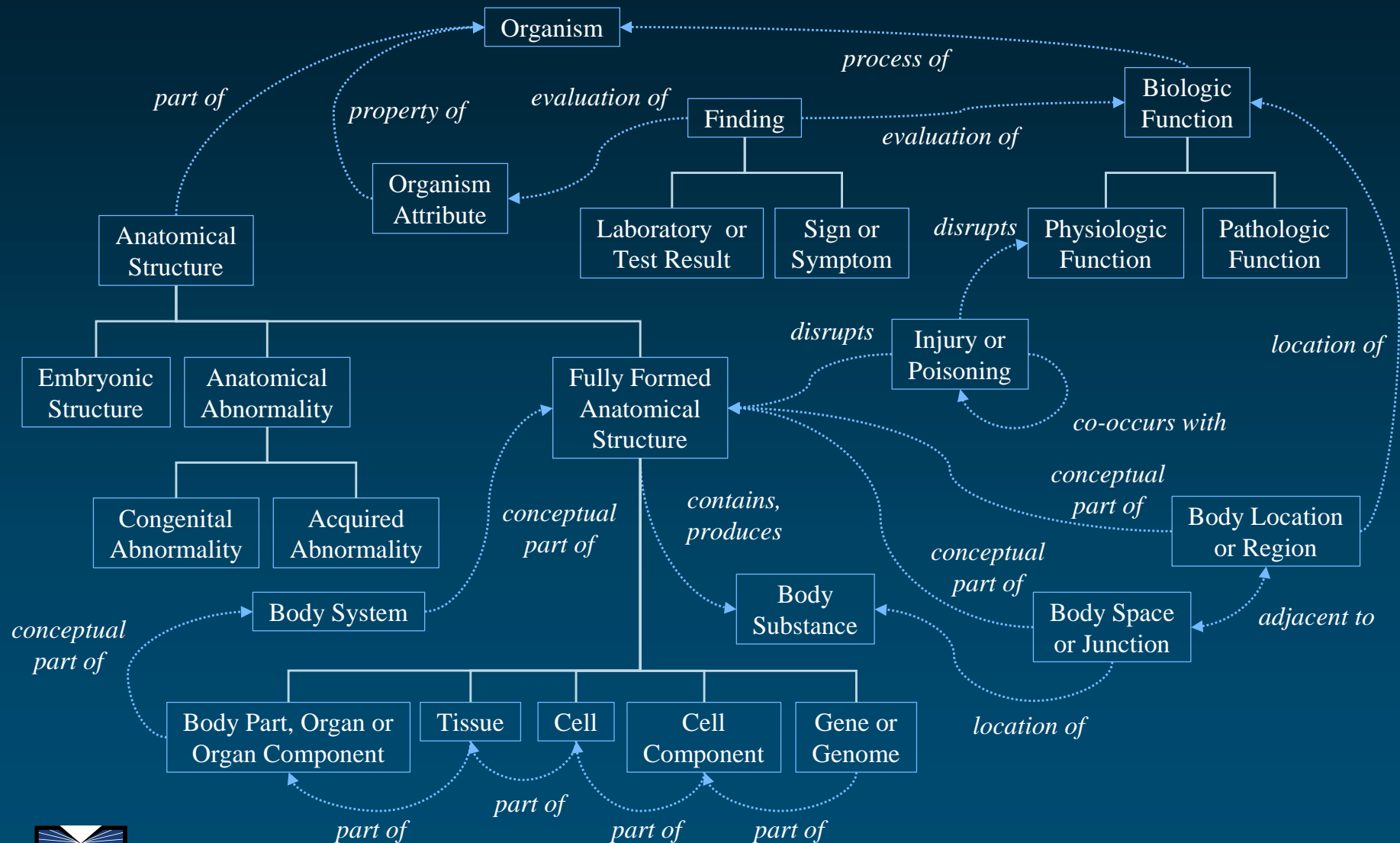
---

- ◆ Semantic network relationships (54)
  - hierarchical (isa = is a kind of)
    - among types
      - **Animal** *isa* **Organism**
      - **Enzyme** *isa* **Biologically Active Substance**
    - among relations
      - *treats isa affects*
  - non-hierarchical
    - **Sign or Symptom** *diagnoses* **Pathologic Function**
    - **Pharmacologic Substance** *treats* **Pathologic Function**

# “Biologic Function” hierarchy (isa)



# Associative (non-isa) relationships



# Role

---

- ◆ A relationship between 2 STs is a possible link between 2 concepts that have been assigned to those STs
  - The relationship may or may not hold at the concept level
  - Other relationships may apply at the concept level
- ◆ A child ST inherits properties from its parents (isa relationships)

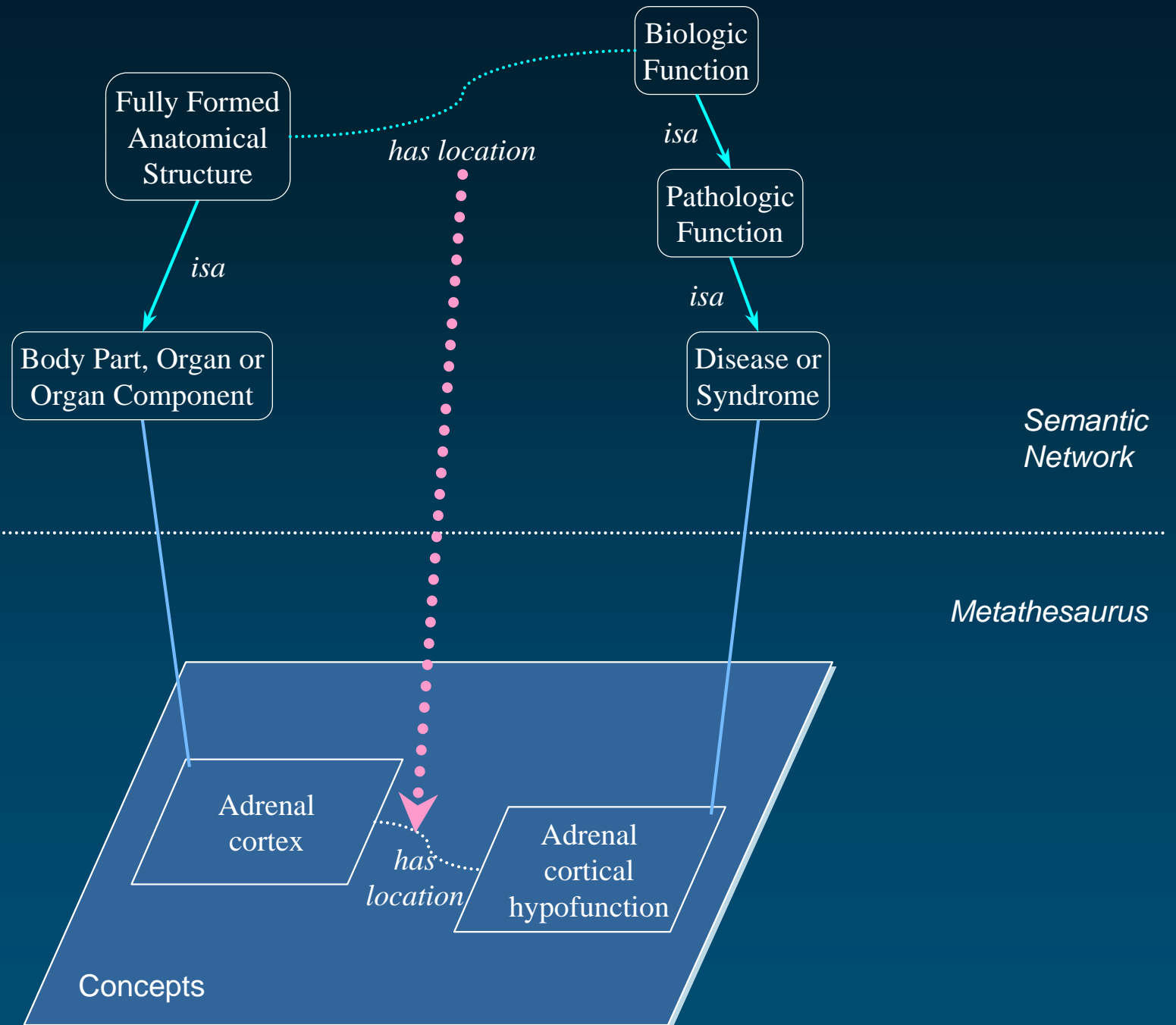
# Applications

---

- ◆ To help qualify inter-concept relationships
  - using the relationships defined between their semantic types in the semantic network
- ◆ To strengthen the structure of the Metathesaurus
  - a relationship between 2 concepts should be consistent with the relationships defined between their semantic types in the semantic network
- ◆ Semantic interpretation
  - finding semantic relationships between concepts in text



## Semantic Types



# Experiment

[McCray & al. (in press)]

- ◆ 3764 concepts related to Heart
- ◆ 6894 pairs of related concepts
  - A relation can be inferred unambiguously from the Semantic Network (65%)
  - Multiple semantic links possible (22%)
  - Violation of the Semantic Network (13%)
    - Wrong inter-concept relationship
    - Wrong categorization
    - Both



# From structure to semantics

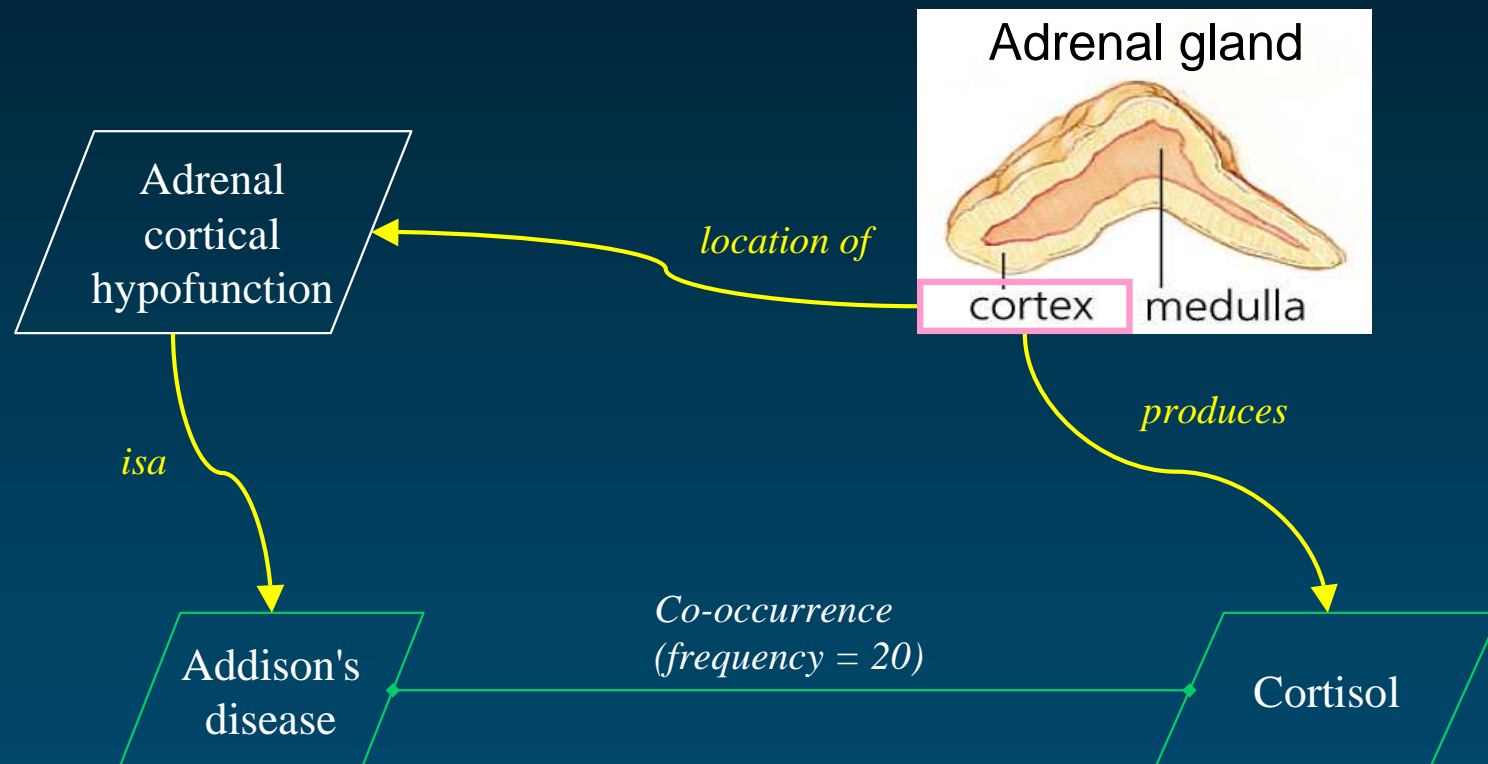
*2. Find a path between two concepts*

# Co-occurrence Overview

---

- ◆ Co-occurrence between MeSH descriptors in MEDLINE citations
- ◆ 8 M pairs of co-occurring concepts
- ◆ Implicit semantics
- ◆ The UMLS provides knowledge for helping make this relationship explicit
  - Corresponding symbolic knowledge (Metathesaurus)
  - Categorization (Semantic Network)

# Co-occurrence Example



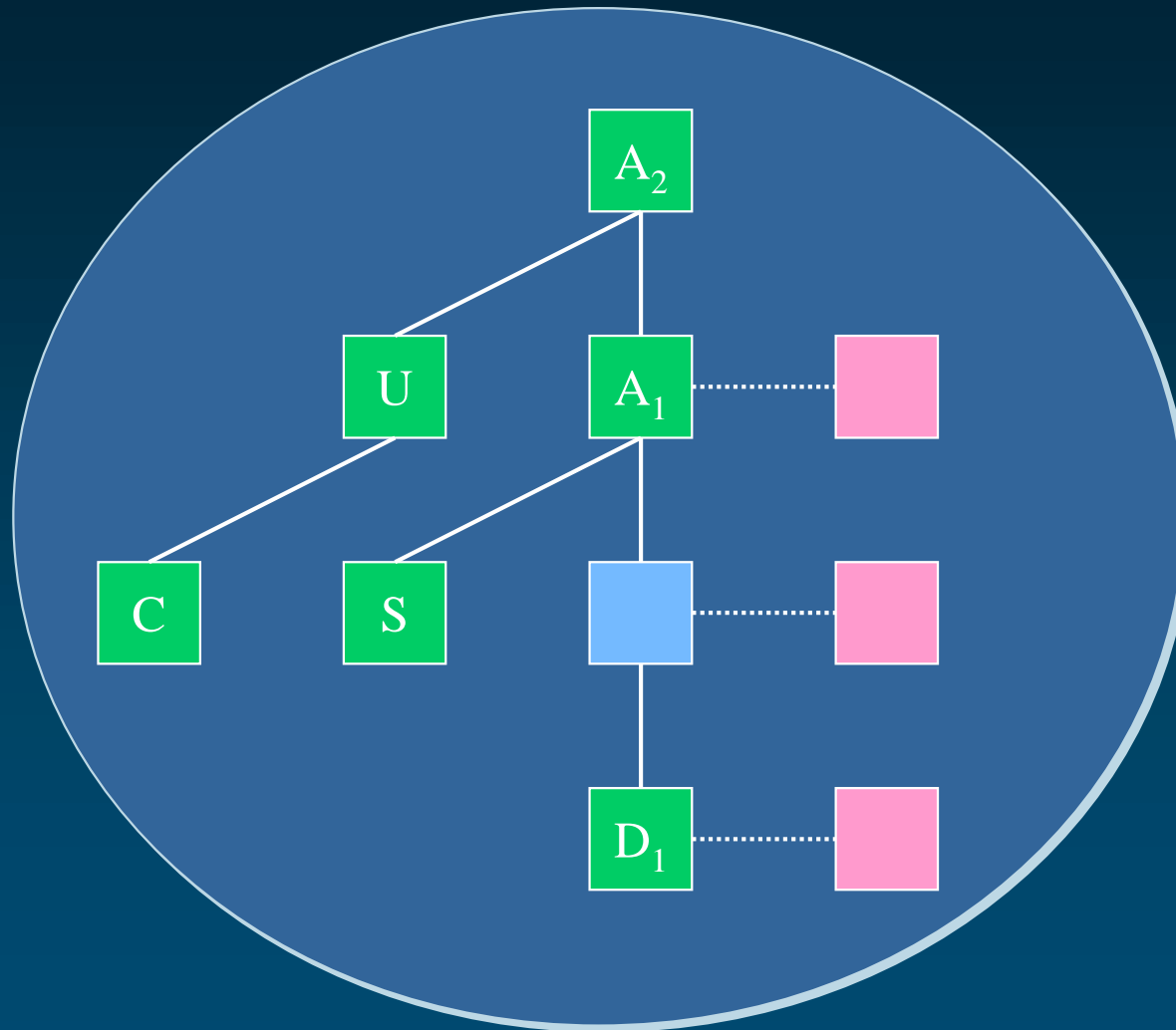
# Co-occurrence Methods

---

- ◆ Based on Metathesaurus relationships
  - Does “Cortisol” belong to the family of “Addison’s disease”?
- ◆ Based on Semantic Network relationships
  - What is the relationship between the semantic types of “Cortisol” and “Addison’s disease”?

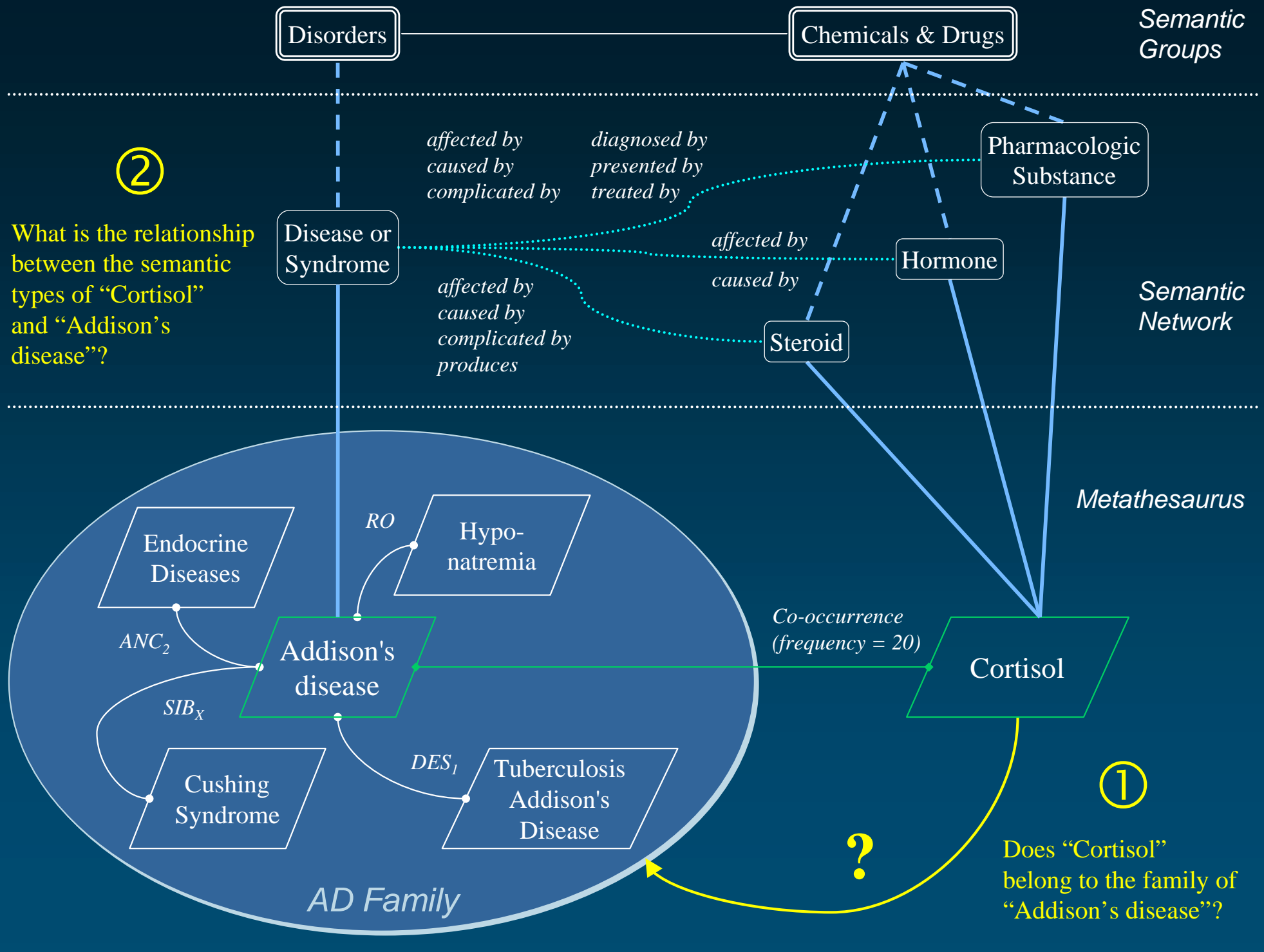


# Family relationships



②

What is the relationship between the semantic types of “Cortisol” and “Addison’s disease”?





# Co-occurrence Results

---

## ◆ Family

- Only 6% of the relationships between co-occurring concepts correspond to symbolic relationships recorded in the Metathesaurus

## ◆ Semantic types

- The semantics of the relationship often remains ambiguous

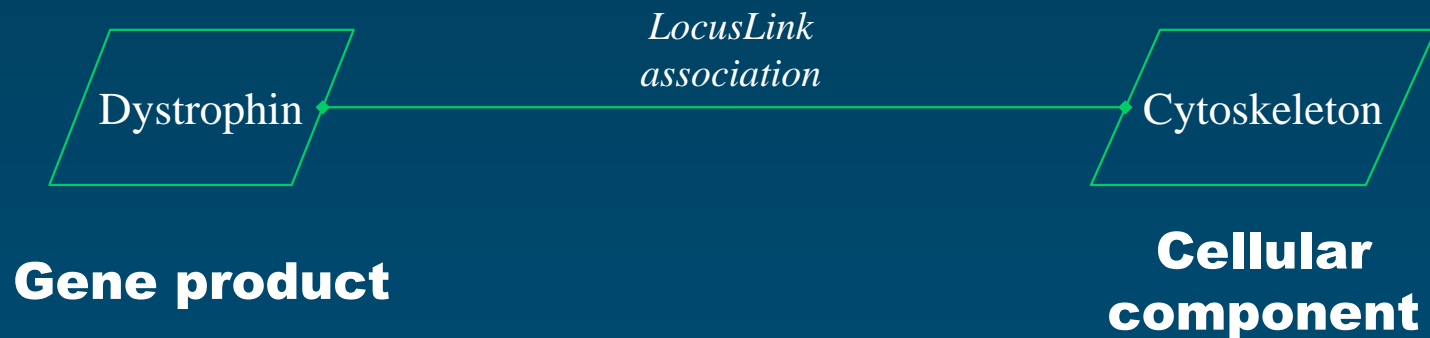
# Bioinformatics Overview

---

- ◆ Association in LocusLink between a Gene / Gene product and
  - Phenotype
  - Molecular function
  - Biological process
  - Cellular component
- ◆ Explicit relationship
- ◆ Most concepts presents in the UMLS
  
- ◆ Is the relationship present in the UMLS?

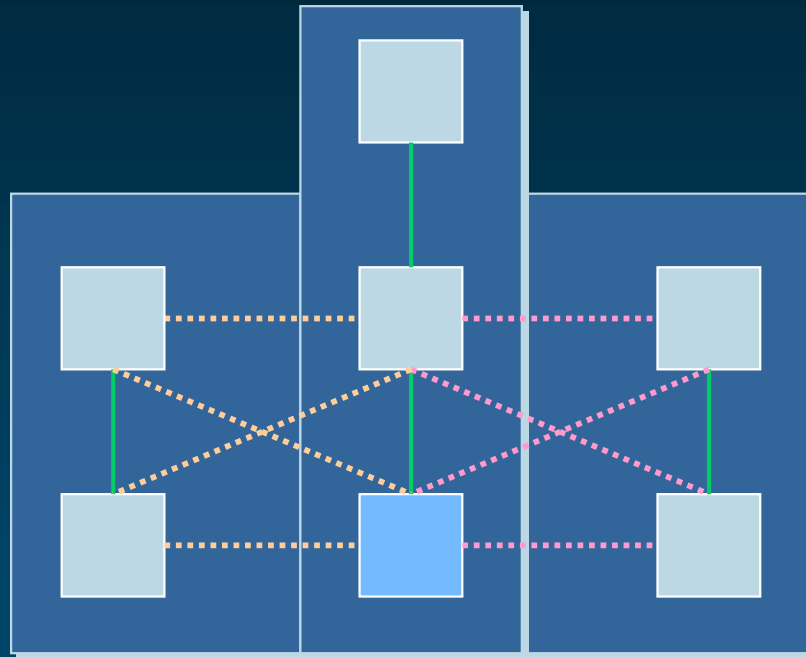
# Bioinformatics Example

---



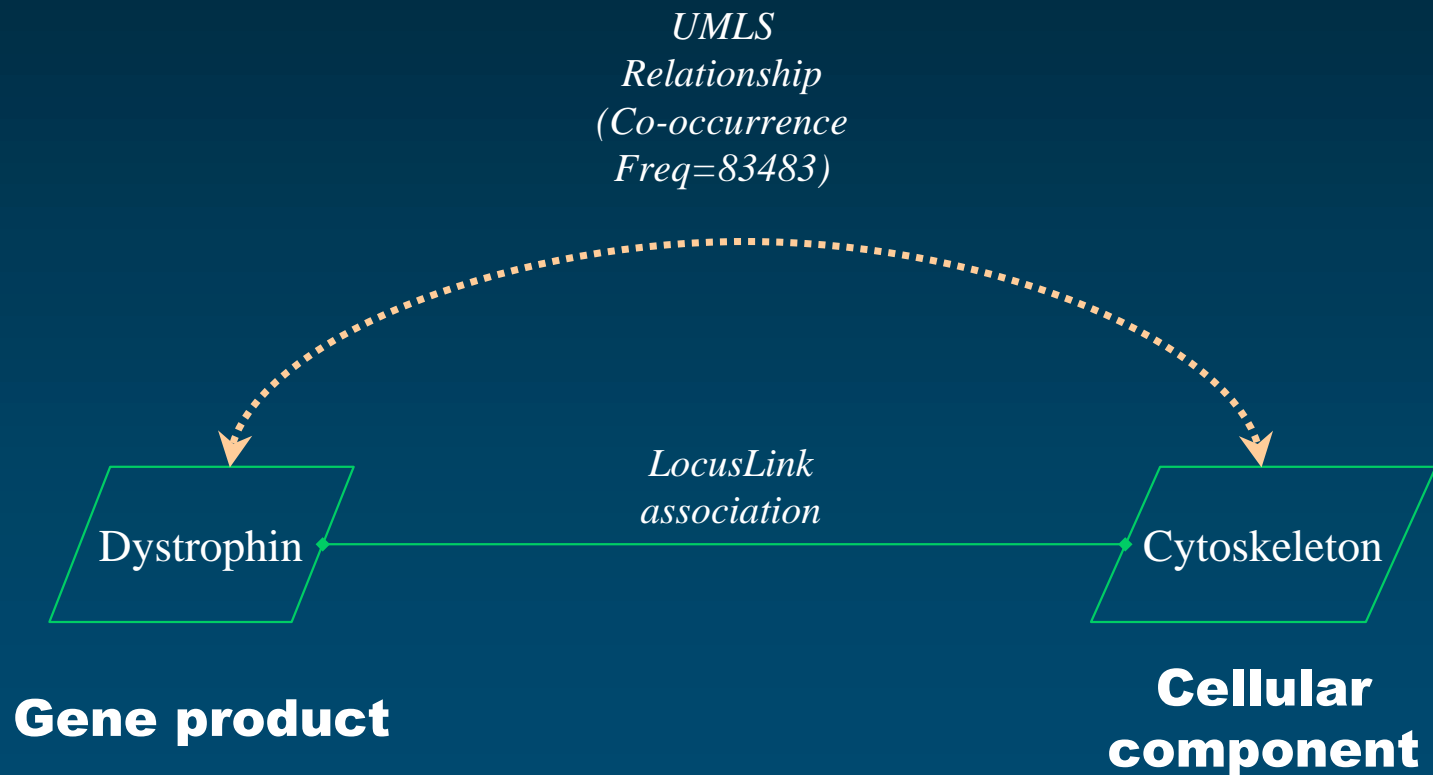
# Bioinformatics Relationships explored

# Hierarchical Associative Co-occurrence



# Bioinformatics Example

---



# Bioinformatics Results

---

- ◆ 70% of LocusLink associations supported by some kind of relationship in the UMLS
- ◆ Many UMLS relationships supporting LocusLink associations are co-occurrence relationships
- ◆ Variation per domain
  - Phenotype 64%
  - Molecular function 83%
  - Biological process 60%
  - Cellular component 70%

# From structure to semantics

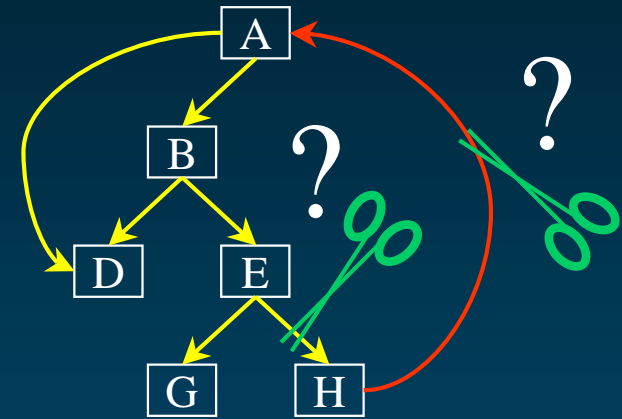
## *3. Limitations*

# Insufficient

## ◆ Identifying vs. solving problems

### ● Cycles

- Which edge to remove?
- Underlying issues



### ● Inconsistency between Semantic Network and Metathesaurus relationships

- Wrong Metathesaurus relationship
- Wrong / missing Semantic Network relationship
- Wrong categorization



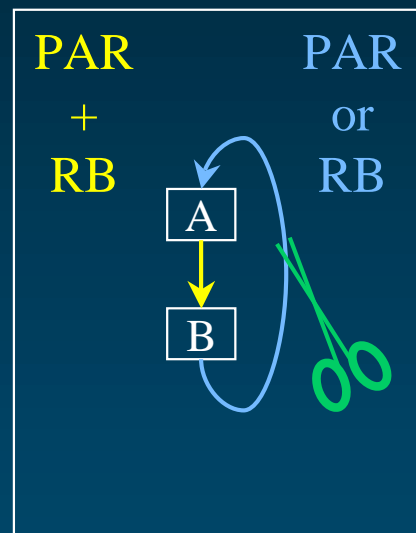
# Insufficient, but...

---

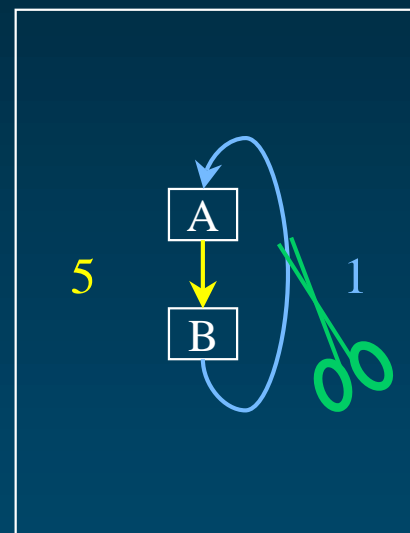
- ◆ Identifying problems is already important
- ◆ Possible uses
  - Retrospectively:  
To focus the work of human editors of the UMLS
  - Prospectively:  
Structural constraints could be used as filters integrated to the UMLS editing environment
- ◆ Additional clues are sometimes available
  - Redundancy
  - Linguistic features

# Structure + redundancy (1)

---

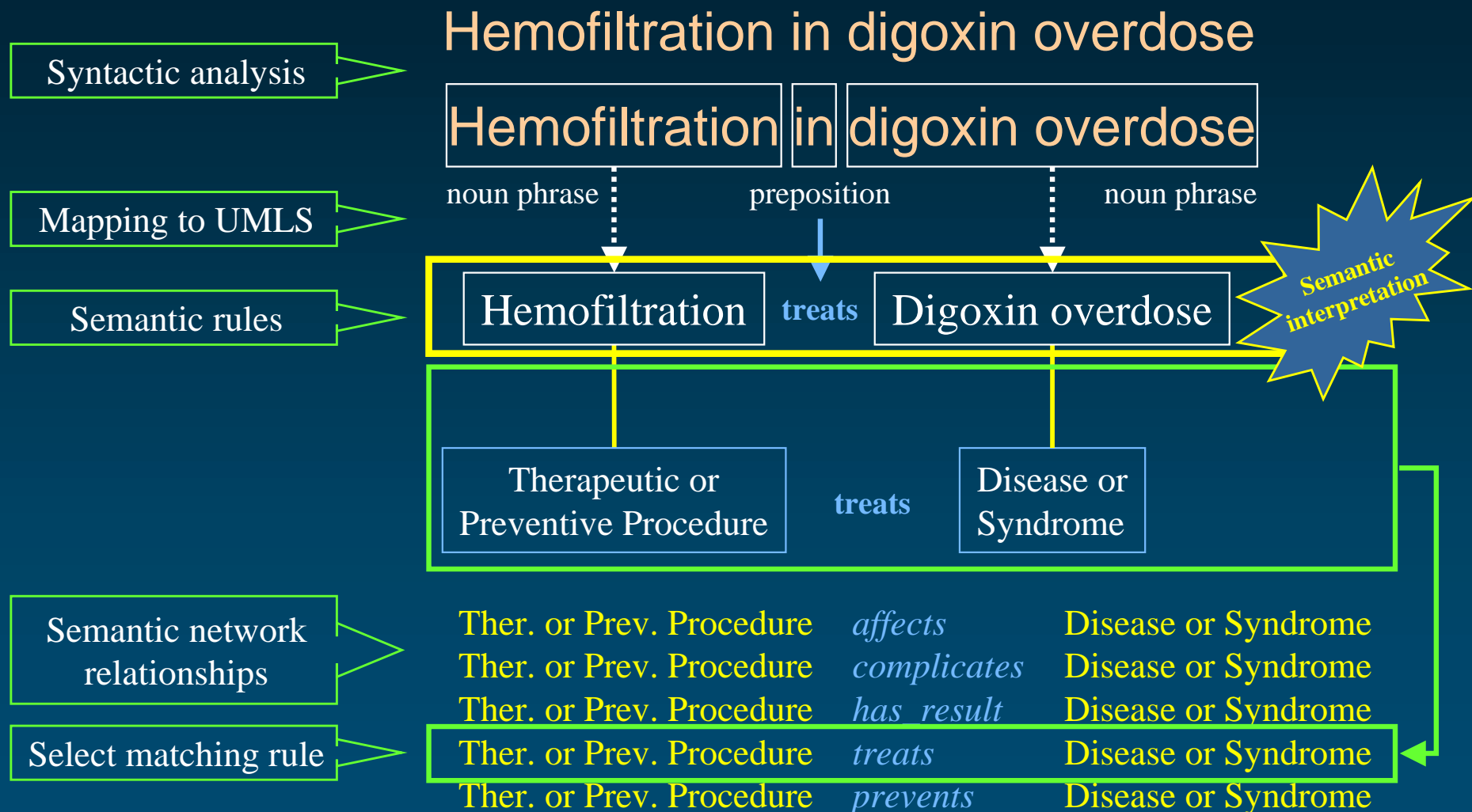


Redundancy



Democracy

# Structure + linguistic features (2)

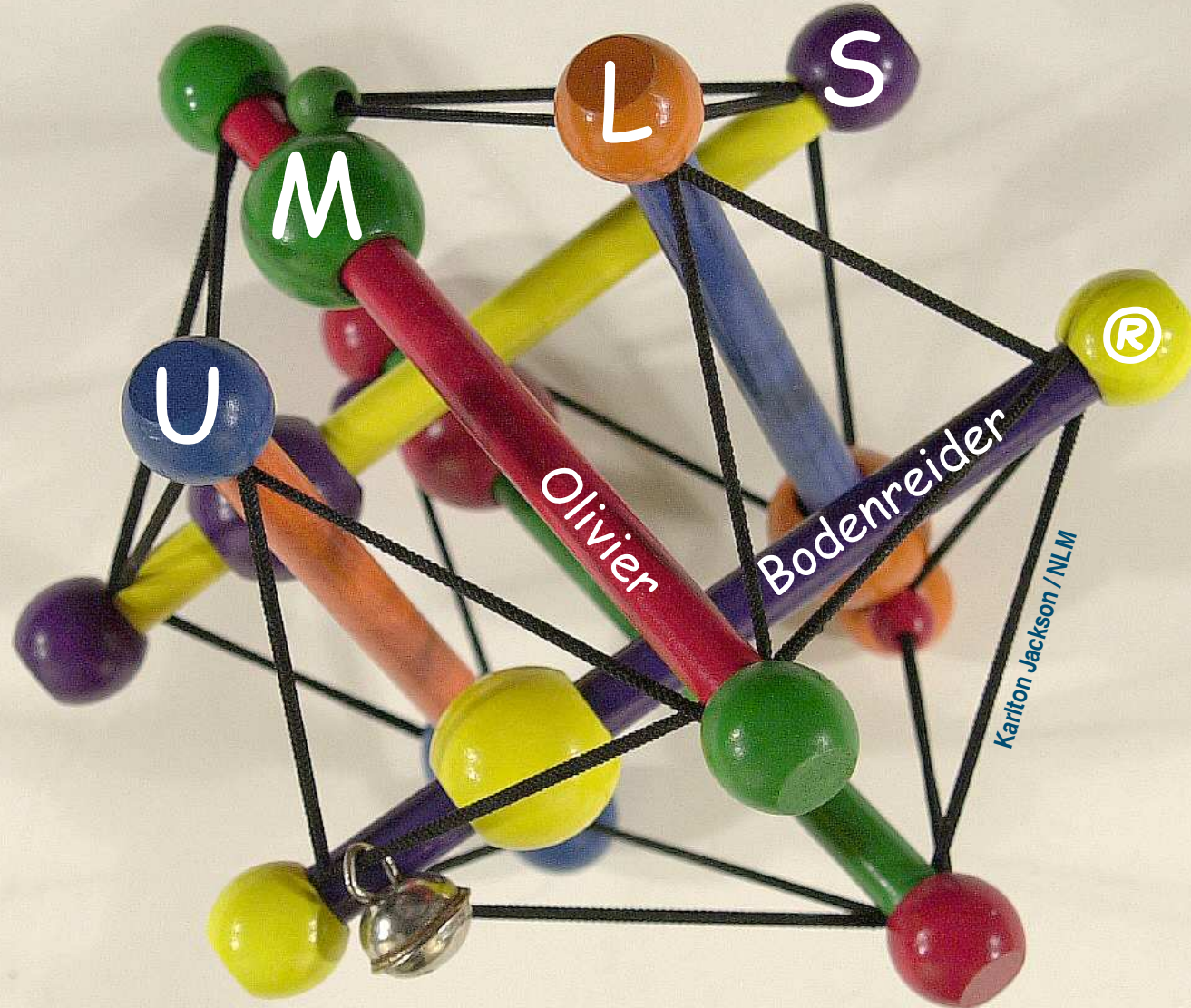


# Conclusions

# Conclusions

---

- ◆ Good knowledge resources should be structurally sound
- ◆ Identifying structural abnormalities may help identify semantic problems
- ◆ Like syntax, structure alone does not ensure that the semantics is correct
- ◆ Close to approaches based on description logics



Karlon Jackson / NLM

A Semantic Space For Kids To Play With<sup>®</sup>



Contact: [olivier@nlm.nih.gov](mailto:olivier@nlm.nih.gov)



*Olivier Bodenreider*

Lister Hill National Center  
for Biomedical Communications  
Bethesda, Maryland - USA